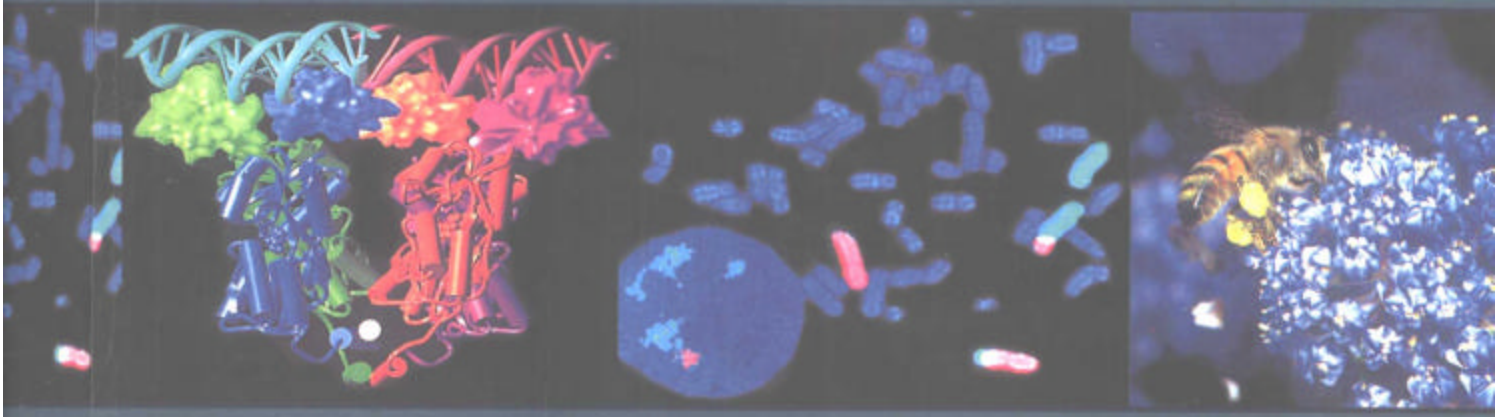
高等院校

高·等·院·校·选·用·教·

现代遗传学原理



徐晋麟 徐 沁 陈 淳 编著

书馆

科学出版社

高等院校选用教材

现代遗传学原理

徐晋麟　徐　沁　陈　淳　编著

Q3-13
XJL

科学出版社

2001

A1C018613379

内 容 简 介

遗传学是生命科学中一门重要的基础与前沿学科，它发展快、应用性强。本书以丰富的资料、图文并茂的形式，系统而详尽地论述了现代遗传学的基础与进展。内容包括遗传信息的传递规律及其物质基础、遗传信息的表达、遗传信息的改变、发育遗传学和分子进化4部分共25章，全面阐述了现代遗传学的所有方面，并附有一定量精心设计的习题和答案，书后还有名词解释、遗传学大事年表等实用资料和索引等。

本书是在多年成功教学的基础上，参考国际权威著作编写而成，其800幅插图可以使读者更好地领会书中的内容。本书特别适合高等院校生物学专业作为本科教材，也可供相关专业的教师和研究生参考。

图书在版编目(CIP)数据

现代遗传学原理/徐晋麟等编著．-北京：科学出版社，2001.3
(高等院校选用教材)
ISBN 7-03-008381-4

Ⅰ.现… Ⅱ.徐… Ⅲ.遗传学-高等学校-教材 Ⅳ.Q3

中国版本图书馆CIP数据核字（2000）第62686号

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
新蕾印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2001年3月第 一 版 开本：850×1168 1/16
2001年3月第一次印刷 印张：60 1/2
印数：1—3 500 字数：1 420 000
定价：68.00元
(如有印装质量问题，我社负责调换〈环伟〉)

目　录

前　言
绪　论 …… 1

第一部分　遗传信息的传递规律和物质基础

第一章　孟德尔定律 …… 7
第一节　孟德尔实验 …… 7
第二节　孟德尔定律 …… 9
第三节　统计学原理在遗传学中的应用 …… 15
第二章　遗传的染色体学说 …… 23
第一节　有丝分裂和减数分裂 …… 23
第二节　在生活周史中的减数分裂 …… 33
第三节　基因和染色体的关系 …… 36
第四节　性染色体的发现 …… 39
第五节　性连锁的发现 …… 40
第六节　染色体学说的直接证据 …… 42
第三章　性别决定和性相关遗传 …… 44
第一节　生物的性别决定 …… 44
第二节　外界条件对性别发育的影响 …… 48
第三节　性相关遗传 …… 49
第四节　人类的性染色体异常 …… 55
第四章　孟德尔分析的扩展 …… 57
第一节　环境的影响和基因表型效应 …… 57
第二节　显隐性关系的扩展 …… 61
第三节　复等位基因 …… 64
第四节　基因间的相互作用与修饰 …… 68
第五章　基因的连锁与交换和真核生物的基因作图 …… 75
第一节　基因连锁的发现 …… 75
第二节　基因重组和染色体交换的作用 …… 77
第三节　交换和重组值 …… 79
第四节　基因定位和遗传作图 …… 80
第五节　真菌的遗传分析 …… 83
第六节　人类染色体作图 …… 95
第六章　数量遗传 …… 97
第一节　数量性状的特点 …… 97
第二节　分析数量性状的基本统计方法 …… 102
第三节　遗传力 …… 108
第四节　人类的数量性状 …… 113
第七章　基因的结构和功能 …… 115
第一节　一基因一酶学说 …… 115
第二节　人类酶缺陷的遗传基础 …… 120
第三节　遗传咨询和产前诊断 …… 124
第四节　基因的精细结构和顺反测验 …… 130
第八章　细菌和噬菌体的遗传重组 …… 133
第一节　基因和基因产物的符号 …… 135
第二节　细菌转化 …… 133
第三节　细菌的接合 …… 137
第四节　高频重组品系 …… 143
第五节　噬菌体染色体的连锁和交换 …… 152
第六节　转导 …… 158
第九章　遗传物质的分子结构和性质 …… 164
第一节　核酸是遗传物质 …… 164
第二节　DNA 和 RNA 的化学组成及双螺旋模型 …… 169
第三节　DNA 的结构和性质 …… 174
第四节　DNA 的变性与复性 …… 180
第五节　分子杂交 …… 183
第十章　基因组和染色体 …… 188
第一节　原核生物基因组的结构特点 …… 188
第二节　真核基因组的复杂性 …… 194
第三节　着丝粒和端粒 …… 203
第四节　染色体的核型和分带 …… 208
第五节　单一序列和重复序列 …… 219
第六节　C 值矛盾 …… 226
第七节　断裂基因 …… 226
第十一章　DNA 的复制 …… 233
第一节　半保留复制的验证 …… 233
第二节　复制的起点、方向和终点 …… 238
第三节　DNA 复制突变型的筛选 …… 242
第四节　原核生物复制的酶系统 …… 245
第五节　原核生物及噬菌体和病毒的复制起始和终止 …… 255
第六节　复制的过程 …… 273
第七节　真核生物的 DNA 复制 …… 279

第二部分　遗传信息的表达

第十二章　转录 …… 293
第一节　信使的发现 …… 294
第二节　RNA 的酶促合成 …… 296
第三节　细菌基因的转录 …… 298
第四节　真核生物基因的转录 …… 307
第十三章　转录后加工 …… 321
第一节　tRNA 和 rRNA 的加工 …… 321
第二节　前体 mRNA 的加工 …… 330
第三节　内含子的剪接 …… 336
第四节　植物病毒和类病毒的自我剪切 …… 353
第五节　编辑 …… 355
第十四章　遗传密码和遗传信息的翻译系统 …… 360
第一节　遗传密码的破译 …… 360
第二节　遗传密码的证实和特点 …… 370
第三节　tRNA 的结构和功能 …… 375
第四节　核糖体的结构和功能 …… 382
第十五章　蛋白质的合成,转运和加工 …… 389
第一节　氨基酰－tRNA 的形成及核糖体的作用位点 …… 390
第二节　肽链的合成 …… 399
第三节　肽链合成的延伸和终止 …… 412
第四节　真核生物蛋白质的越膜运输 …… 419
第五节　细菌蛋白的越膜运输 …… 433
第六节　翻译后的加工 …… 436
第十六章　原核生物基因表达的调控 …… 442
第一节　转录水平的调控 …… 443
第二节　操纵子的其他调控形式 …… 466
第三节　转录的终止调控 …… 471
第四节　RNA 聚合酶转录起始 …… 475
第五节　DNA 重排对转录起始的调控 …… 484
第六节　翻译的调控 …… 486
第十七章　噬菌体的裂解途径和溶原化途径的选择和调控 …… 503
第一节　λ 噬菌体的基因组及早、晚期转录 …… 504
第二节　溶原化途径的维持 …… 512
第三节　阻遏物合成的建立 …… 521
第四节　Cro 蛋白的功能 …… 524
第五节　溶原化和裂解周期的平衡 …… 526
第十八章　真核生物基因表达的调控 …… 529
第一节　染色体水平的调控 …… 530
第二节　真核生物中染色体的重排 …… 537
第三节　染色质水平的调控 …… 563
第四节　DNA 水平的调控 …… 571
第五节　转录水平的调控 …… 573
第六节　转录起始和加工的调节 …… 581
第七节　翻译的调控 …… 583
第八节　细胞周期的调控 …… 587
第十九章　癌基因与癌 …… 592
第一节　肿瘤细胞的特点 …… 592
第二节　癌基因和抗癌基因 …… 594
第三节　癌基因产物的结构、功能和细胞癌变 …… 602
第四节　染色体畸变和肿瘤的发生 …… 610
第二十章　核外遗传信息的传递 …… 615
第一节　核外基因组的结构 …… 615
第二节　核外遗传 …… 623
第三节　感染遗传——杀伤性酵母 …… 632
第四节　母体影响 …… 633
第五节　真核细胞中的质粒 …… 635

第三部分　遗传信息的改变

第二十一章　基因突变 …… 639
第一节　基因突变的类型 …… 640
第二节　突变的原因 …… 641
第三节　突变和人类疾病 …… 652
第四节　DNA 的修复机制 …… 655
第二十二章　染色体畸变 …… 665
第一节　断裂愈合和交换学说 …… 666
第二节　染色体结构的改变 …… 667
第三节　染色体畸变的鉴别 …… 681
第四节　染色体数目的改变 …… 684
第二十三章　重组和转座 …… 696
第一节　重组机制 …… 698
第二节　Holliday 模型 …… 701
第三节　双链断裂起始重组 …… 705
第四节　重组和酶 …… 709
第五节　位点特异性重组 …… 714
第六节　原核生物的转座因子 …… 718
第七节　真核生物的转座因子 …… 733

第八节 反转录病毒和反转录子 …… 744

第四部分 发生遗传学和分子进化

第二十四章 基因调控和细胞的分化发育 …… 765
第一节 高等生物的细胞核是否有全能性 …… 766
第二节 线虫的发育模式 …… 771
第三节 果蝇的早期发育 …… 773
第四节 母体基因的调节作用 …… 779
第五节 分节基因的调节作用 …… 788
第六节 同源异型复合基因座 …… 795
第七节 动物的性别决定和分化 …… 804
第二十五章 群体遗传学 …… 816
第一节 基因型频率和基因频率 …… 817
第二节 哈迪－温伯格法则 …… 820
第三节 在自然群体中的遗传变异 …… 826
第四节 群体中基因频率的改变 …… 830
第五节 分子进化 …… 848

习题(附参考答案) …… 857
参考文献 …… 889

附 录

名词解释 …… 901
遗传学领域诺贝尔奖名录 …… 925
遗传学大事年表 …… 927
果蝇和玉米的遗传图 …… 933

索 引 …… 935

前　言

自国内开设遗传学课程以来，诸多老一辈遗传学家编写了很多优秀的遗传学教材，其中方宗熙先生编写的《细胞遗传学》(科学出版社，1961 年，后改为《普通遗传学》)和刘祖洞、江绍惠先生编写的《遗传学》(高等教育出版社，1981 年)最有影响，培养了一代又一代人才，我本人也受益匪浅。

近 20 年遗传学发展迅猛，国内传统教材已难以适应现代教学的要求。在长期从事遗传学教学和科研中，我感到迫切需要一部新的教材，使读者能从细胞和分子两个不同的层次了解基因纵向传递和横向传递的规律与机制。此教材应做到内容新颖，能及时反映遗传学的现状和进展；插图应有自己的特色，图文并茂；书后应附有参考文献、索引、名词解释、习题和参考答案以及图表的出处等。这样使读者阅读和追踪都比较方便。我自己一直很想在这方面做些尝试，但考虑到巨大的工作量而不敢轻举。

两年前，我和徐沁、陈淳两位博士谈了这一想法，大家都有同感，于是决定分工合作，来编写一本能适应遗传学发展的新教材，在内容上将经典遗传学和分子遗传学有机地融合在一起，特别注重介绍遗传学的最新进展，详细介绍重大发现的思维过程和科学家们前仆后继的奋斗精神；在形式上大量采用插图，以帮助读者更好地理解。由于我们水平有限，本书距离我心目中理想的教材尚有较大的差距。希望我们的努力对遗传学的研习者能有一定的参考作用，并起到抛砖引玉之效。成稿仓促，谬误之处难免，望读者指正，以便再版时更正。

为了和传统的遗传学教材相区别，又不和 Francisco J 等编写的 *Modern Genetics* 及 Gardner EJ 等编写的 *Principles of Genetics* 重名，遂将本书命名为《现代遗传学原理》(*Principles of Modern Genetics*)。本书主要参考了以下几本名著：①Griffiths AJF, Miller JH, Suzuki DT, Lewontin RC, Gelbart WM. 1993. *An Introduction to Genetic Analysis*, 5th ed., W. H. Freeman and Company, New York；② Russell PJ. 1992. *Genetics*, 3rd ed. HarperCollins Publishers；③Lewin B. 1997. *Genes* Ⅵ. Oxford University Press, Oxford。

本书篇幅较大，出版周期短，科学出版社编辑马学海博士付出了很多的精力和辛劳。在编写过程中姜勇和张勇两位同学在文稿录入和部分习题的编译中做了很多工作，在此一并致谢。

徐晋麟

2000 年 9 月于上海交通大学

绪　论

遗传学是生命科学领域中一门核心学科，主要是研究遗传物质的结构与功能以及遗传信息的传递与表达。这一学科名称是由英国遗传学家贝特森(Bateson W)于1909年首先提出的。经典遗传学主要是研究遗传物质纵向传递的规律以及表型和基因的关系。分子遗传学则偏重研究基因的结构、功能和横向传递。结构是指其化学本质与精细结构，功能是指遗传物质的复制、表达、调控、重组与变异。而群体遗传学是研究群体中基因频率和基因型频率以及影响其平衡的各种因素。

遗传学和生命科学其他学科相比，有着自己鲜明的特点：①是一门推理性的学科，而不是描述性的。研究遗传学的方法很像物理学，是根据自然现象或实验的数据推理出一种假说，然后再通过实验来加以验证。②多学科的交叉和融合。遗传学主要是建筑在生物化学、细胞生物学和统计学三门学科上，但又涉及到生命学科的各个领域，如动物学、植物学、微生物学、医学、农学、免疫学、系统分类学、生理生态、代谢调控等，甚至还涉及一些社会科学，如心理学、社会学、犯罪学等。③发展快。遗传学的发展非常快，可以说是“月新年异”，新理论、新技术、新成果层出不穷，几年前出版的书籍常常已经“过时”，“多利”这头克隆羊的诞生一下子就把动物细胞核不具全能性的理论彻底打破，这样爆炸性的例子在遗传学中也并不少见。④应用性强，转化为生产力的周期短。1953年Watson和Crick提出DNA双螺旋模型时，人们并不知它有什么实际应用价值，但到了20世纪70年代就出现了体外重组技术，现在基因工程技术已成为各国的支柱产业之一，没有双螺旋模型就不可能有重组技术。以遗传学为理论基础，又不断派生出许多应用学科，如动植物及微生物育种学、优生学、生物工程等。

第一节　遗传学的发展史

一、在孟德尔以前及同时代的一些遗传学说

公元前5世纪希波克拉底(Hippocrates)提出了第一个为人所知的遗传理论，他认为子代之所以具有亲代的特性是因为在精液或胚胎里集中了来自身体各部分的微小代表元素，按照这一观点，他相信后天获得的性状是遗传的。100年后，亚里士多德(Aristotle)认为亲代残缺，下一代并不残缺。因而提出，精液不是提供胚胎组成的元素，而是提供后代的蓝图。生物的遗传不是通过身体各部分样本的传递，而是个体胚胎发育所需的信息传递，可惜的是这一精辟而深刻的见解未能引起人们的重视，而一些荒诞的推测却使人们产生了兴趣，如长颈鹿是骆驼和豹子的杂种后代等。

1809年伟大的生物学家拉马克(Lamarck，JB)提出了**“用进废退”**的进化论观点，由此而得出获得性状(acquired characteristics)是可以遗传的。可以说这一错误的观点是他一生中最大的一次失误，可悲的是此观点一直延续到19世纪60年代。但瑕不掩瑜，拉马克在生命科学中的丰功伟绩，使人们公认他是自然科学史中最伟大的科学家之一。

1866年达尔文(Darwin)提出了**泛生论**(hypothesis of pangenesis)，认为身体各部分

细胞里都存在一种胚芽或“泛子”(pangens)，它决定所在细胞的分化和发育。各种泛子随着血液循环汇集到生殖细胞中。受精卵发育过程中，泛子又不断地流到不同的细胞中，控制所在细胞的分化，产生一定的组织器官，显然这一观点是受到希波克拉底的理论的影响，似乎言之有理，但在血液中根本就找不到所谓的“泛子”，所以这一理论是不能成立的。

1883年法国动物学家鲁·威廉(Roux，W)提出有丝分裂和减数分裂过程的存在可能是由于染色体组成了遗传物质，同时他还假定了遗传单位沿着染色体丝作直线排列，当时他并不知道孟德尔已证实了这种遗传单位的存在。

德国的生物学家魏斯曼(Weismann，A)做了连续22代剪断小鼠尾巴的实验，方法虽然简单，但有力地否定了泛生论。1883和1885年他将Roux理论发展成为完整的遗传和发育的理论——**种质论**(germplasm theory)，认为多细胞生物可分为**种质**(germ plasm)和**体质**(somatoplasm)两部分，种质是独立的、连续的、能产生后代的种质和体质。体质是不连续的、不能产生种质。种质的变异将导致遗传的变异，而环境引起的体质的变异是不连续的。他还假设遗传物质在生殖细胞中数量减半，受精时得到恢复。个体的遗传物质一半来自父本，一半来自母本，这一推理是何等的正确！遗憾的是由于历史的局限，魏斯曼误认为细胞核中的每一条染色体都带有个体的全部遗传因子。

1869年达尔文的表弟高尔顿(Galton，F)用数理统计的方法研究人类智力的遗传，发表了“天才遗传”(hereditary genius)，认为变异是连续的，亲代的遗传性在子女中各占一半，并彻底混合，即**“融合遗传论”**。由于他所选择的研究性状是数量性状，所以虽然他的结论是完全正确的，但只适合数量性状，而不能作为遗传的普遍规律。

二、遗传学的诞生

在孟德尔之前已经有一些植物学家做了植物杂交实验，并获得了显著的成绩。如1797年英国的奈特(Knight，T)将灰色和白色的豌豆进行杂交，结果杂交一代全部是灰色，杂交第二代却产生灰色和白色两种，但奈特并未进一步地统计分析，只发现了这一现象。就在孟德尔理论发表的前两年(1863年)，法国的诺丹(Nauding)发表了植物杂交的论文并获法国政府的奖励，他认为：①植物杂交的正交和反交结果是相同的；②在杂种植物的生殖细胞形成时“负责遗传性状的要素互相分开，进入不同的性细胞中，否则就无法解释杂种二代所得到的结果。”这一结论和孟德尔定律已经非常接近，说明孟德尔的发现并非偶然，也是在前人辛勤工作的基础上建立起来的。大部分重大的科学发现都是这样通过几代人的研究、积累、充实、修正而最终得以完善的。

1865年，当时属奥地利的布隆(Brünn)基督教修道院的修道士格里高·孟德尔(Gregor Johann Mendel)，根据他8年植物杂交实验的结果，2月8日在当地的科学协会上宣读了一篇题为“植物杂交实验”的论文，1866年正式发表在该协会的会刊上，并将论文分寄给当时一些国际著名的生物科学家，大概当时因作者的平凡和杂志的低档，孟德尔几乎得不到回音，这一伟大的发现一直被埋没在旧纸堆中长达35年。孟德尔临终前说：“等着瞧吧，我的时代总有一天要来临。”果然这一天终于来临，1900年荷兰的狄夫瑞斯(Devries)，德国的植物学家科伦斯(Correns，CE)和奥地利的切尔迈克(Tschermak)，分别同时发现了这篇论文和他的价值，不论三位学者出

于什么目的，他们最终使孟德尔的学说重见天日，并建立了遗传学这门学科，应当说功不可没，这就是孟德尔定律的二次发现。

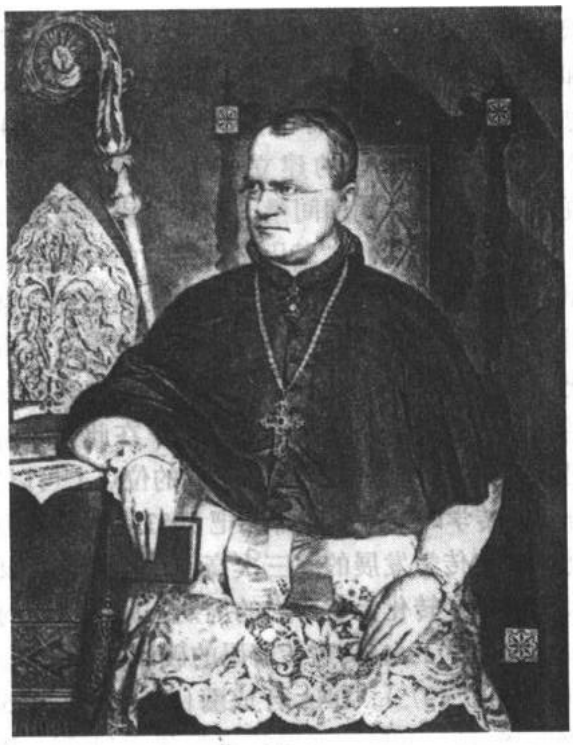


图 1 G. 孟德尔(Gregor Mendel)(1822～1884)。(转引自 Griffiths, A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 1-1)

接踵而来的就是一场激烈的论战：一方是以牛津大学动物学教授韦尔登(Weldon)为首，推崇高尔顿学说，贬低孟德尔学说的贡献，由于他的影响极大，《自然》(*Nature*)等杂志都不发表孟德尔观点的文章；另一方的主将是剑桥大学的遗传学教授贝特森，但由于力量单薄，只得靠私人印发自己的论文来应战。直到 1904 年贝特森在论战中获得胜利，孟德尔主义才摆脱了冷嘲热讽和忽视。1901 年 Devries 提出“突变”这一名词，1902 年 W.S.Sutton 等提出了遗传的染色体学说，1902～1909 年贝特森先后创用**遗传学**(genetics)、**等位基因**(allele)、**纯合体**(homozygous)、**杂合体**(heterozygous)、**上位基因**(epistatic genes)等名词。1909 年丹麦的科学家约翰逊(Johannsen)创用了**基因**(gene)、**基因型**(genotype)和**表型**(phenotype)。此时遗传学的雏形已形成，两大定律已建立，孟德尔定律的广泛适用性已得到了承认，作为一门新的学科终于诞生了。

三、 遗传学的发展

遗传学的发展大致可以分为三个时期。

第一个时期是细胞遗传学时期，约从 1910～1940 年。此时期主要是确立了遗传的染色体学说。较为突出的工作是 1910 年摩尔根(Morgan, TH)带领着他的三大弟子斯特蒂文特(Sturtevant)、布里吉斯(Bridges)和缪勒(Muller)创立了连锁定律，并证实

了基因在染色体上以线状排列。经典遗传学的基本单位是一个不可再分而且是抽象的基因。

第二个时期是微生物遗传学和生化遗传学时期(1941～1960)，在这20年的时间里遗传学有着突飞猛进的发展，研究的对象从真核转到了原核，更为深入地研究了基因的精细结构和生化功能。重大成果有“**一基因一酶**”(Beadle and Totum，1941)的建立，遗传物质确定为DNA，而不是蛋白(Avery，1944)；**跳跃基因**的发现(McClintock B，1951)；**双螺旋模型**的建立(Watson 和 Crick)以及**中心法则**的提出(Crick，1958)。值得一提的是在此期间量子物理学家薛定鄂(Schrödinger，1945)出版了《什么是生命？》一书，指出：“基因是活细胞的关键组成部分，要懂得什么是生命就必须知道基因是如何发挥作用的。”这本书向物理学家们预告一个生物学研究的新纪元就将开始，值得大家奋起钻研。很多物理学家都纷纷转向遗传学这个新领域进行研究，把物理学的思维方式也带入其中，促使遗传学的研究方法和思维方式发生了一场大的变革，从而获得了长足的发展。在此期间遗传的基本单位是**顺反子**(cis-trons)，它是具有一定功能的实体，在不同的位点上可以发生突变和重组。

第三时期是分子遗传学时期。一般都会把1953年双螺旋模型的建立作为此期的起点直到现在，此期是遗传学发展的第三次高潮，可以说成果累累，月新年异，而且趋向于应用，大大缩短了转化为生产力的周期。此期主要的贡献是乳糖操纵子模型的建立(Jacob and Monod，1961)，遗传密码的破译(Nirenberg and Khorana，1964，1965)，反转录酶(Temin，1975)，DNA合成酶(Kornberg，1958)，限制性酶的发现(Arber，1962，1968；Smith，1978)，重组技术的建立(1972，Berg)，DNA测序(Sanger and Gilbert，1977)，转座子的移动(Shapiro，1980)，核糖酶的发现(Cech and Altman，1981)，PCR技术的建立(Swithies，1986)，内含子的发现(Sharp and Roberts，1977)，近来克隆羊的成功(Wilmut，1997)以及人体遗传密码草图面世（2000年6月26日宣布）。此期基因的概念是一段可以转录为功能性RNA的DNA，它可以重复、断裂的形式存在，并可转座。

第二节 遗传学在国民经济中的作用

一、遗传学与农牧业的关系

无论是农林还是畜牧水产业都是和国计民生紧密相关的，其中心的问题就是“种”的问题。所谓的优良品种主要体现在产量、质量、抗病害三大指标上。改良品种，甚至创造新的品种，人们都是依据遗传学原理，利用诱变、杂交、细胞工程、基因工程等方法来实现的。如我国的杂交水稻、小黑麦等大大地提高了产量、提高了品质，还以基因工程的方法培养了抗病害的棉花、抗除草剂的作物等。人们还期望把固氮基因转入非豆科的粮食作物中，以节省肥料提高产量。另一途径是培养高光效植物，以充分利用光能。克隆羊的成功给无性繁殖优良家畜带来了曙光。

二、遗传学与工业的关系

和遗传学关系密切的有生物制药、化学工业和食品工业、发酵工业等。人们可以利用遗传原理来进行工业微生物的育种，用基因工程的途径制备各种工程菌，还可以改变酶的分子结构以提高其活性。重组的生物制品现已发展成为一项支柱产业，干扰素、

胰岛素、白细胞介素－2等重组产品已正式投入市场。人们还想把蜘蛛丝蛋白基因克隆出来，用于生产高强度的丝纤维。

三、遗传学在能源的开发和环境保护中的应用

利用工程菌可以水解植物的茎秆，产生乙醇，变废为宝。还可以通过厌氧发酵使工业废水产生沼气；利用工程菌来富积废水中的重金属，不仅节约资源，还可清除污染；还可用于三次采油，以及消除海洋中的原油污染等。利用Ames法、染色体畸变、姐妹染色体交换、微核技术以及果蝇CIB等系列技术等，可以检测致癌、致畸变和致突变物质。

四、遗传学在医疗卫生工作中的应用

人类疾病中存在四大难题：肿瘤、心血管疾病、遗传病和某些病毒感染（如艾滋病、埃伯拉病毒和疯牛病等），这些难题和遗传都紧密相关。肿瘤的本质是癌基因的突变或调控的改变影响其产物的质和量，从而造成细胞内信息传递紊乱所致；心血管疾病中有的也具有遗传性；遗传病已经发现有4000多种，是由基因突变所造成。艾滋病等虽不是人类本身的基因突变所致，但要想获得有效的防治方法，必须首先搞清这些病毒基因组的结构，及其复制和表达的规律，从而针对性地制定防治方法。

基因治疗以及反义技术的应用等离临床应用尚存一段距离，但却给人类带来很大的希望。

当今社会遗传学涉及面已经很广，如法律上亲子鉴定、犯罪嫌疑人的排查、考古中DNA的鉴定、体育中人才的选拔等都或多或少和遗传有关，可以说遗传学是一门事关国计民生、人寿年丰的学科。

第一章 孟德尔定律

基本概念

基因型是一种有机体的内在遗传结构，表型是基因型外在的生理表现。

基因对性状的发育赋予潜在的作用。这种作用受到其他基因的互作和环境因子的影响。

孟德尔的分离法则是指一对等位基因在形成配子时彼此分离。自由组合法则则是指不同染色体上的基因在配子形成时是彼此自由、随机地被组合到子细胞中。

测交是将未知基因型的个体（常存在显性表型）和一个纯合隐性基因型的个体进行杂交来确定基因型。

孟德尔法则适用于人类及一切真核生物。人类性状遗传的研究不能通过杂交，而只有通过谱系分析和其他生化及分子生物学的实验来进行。

第一节 孟德尔实验

一、孟德尔的功绩

孟德尔在前人工作的基础上进行了植物杂交实验，并且一针见血地指出，前人的工作“到目前为止，还没有卓有成效地提出一个能普遍应用的控制杂种形成和发育的规律”。还指出了存在的问题：①没有一个试验能确定杂种后代出现的类型数目；②没有按照不同世代把这些类型予以归类；③没有搞清它们在统计学上的关系。针对这些问题孟德尔开展了 8 年工作，采用 32 个品种，观察了 7 对性状（表1-1），最终总结了两个定律，即独立分配和自由组合定律，后人称为孟德尔定律，从而创建了遗传学。其贡献与当时 19 世纪三大发现相比毫不逊色，遗憾的是不幸被埋没了 35 年，这是遗传学发展史上的一个悲剧。

二、孟德尔植物杂交实验

孟德尔实验的成功归功于他卓越的观察力和方法学。

1. 设计严密，层次分明

(1) 亲本杂交（parent generation） ①选材：孟德尔选材时考虑到“这种植物的杂种在开花期必须防止所有外来花粉的影响”。刚开始，鉴于豆科植物的特殊花器结构成为被选的实验材料，最终选择了严格自花受粉的豌豆；在实验中他又注意选择生命力强的种子，使后代的成活力相同，避免统计误差；种子在使用前进行试种，观察是否存在分离，确定纯种后才使用。② 选择研究的性状：“具有稳定的可

区分性”，即质量性状，摒弃数量性状，抓住了遗传规律的主流。③ 继承了前人采用的**正反交**（reciprocal cross），这为以后区分伴性遗传和细胞质遗传打下了基础。④ 设立对照实验：将盆栽的植物移到温室内和大田种植的作对照，以排除虫媒的干扰。

（2）杂种 1 代（filial generation 1，F_1）的观察　确定了**“显性”**（dominant）性状和**“隐性”**（recessive）性状。将在杂交中能传递给后代的性状称为显性，而在传递过程中潜伏的性状称为隐性。

（3）F_2 代的观察　①对后代的性状进行了统计处理，发现显性性状和隐性性状的分离比为 3∶1；②注意到了“当试验植株数目较少时，结果可发生相当大的波动”，因而采用大样本。③为保证后代成活、统计精确，十分细致地将矮株进行移植，避免其被相邻的高株覆盖因得不到阳光而死亡。④对具有显性性状的植株，进一步通过自交来研究其遗传结构。

（4）F_3 代的研究　经观察统计表型为显性（A）的植株有两种基因型：AA 和 Aa，比例为 1∶2。

2. 科学推论

对以上的结果，孟德尔进行了科学推论：　①遗传因子成对存在，形成生殖细胞时彼此分离；②以 A 代表显性性状，a 代表隐性性状，Aa 表示杂合类型，那么在 F_2 代中后代的构成可用公式：1A + 2Aa + 1a 来表示，表明了各种类型的基因型以及比例。

3. 精确验证

孟德尔不仅推论出公式并创用**测交**（test cross）方法对推论加以验证(图 1－1)，突破了传统生物学的研究方法。测交是指将杂种后代和隐性亲本进行杂交，回交（back cross）是指杂种后代和任一亲本杂交。所以测交是回交的一种，这种方法可以测出杂种后代的基因型，故称为测交。

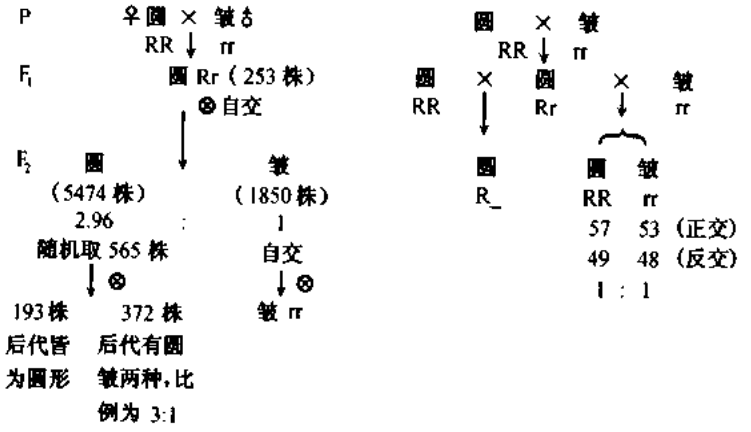


图 1－1　豌豆的杂交和回交实验。

如果推测是正确的话，那么圆形的基因型为 RR，皱缩的基因型为 rr。

表 1-1 孟德尔的豌豆杂交实验 7 对性状的结果

豌豆表型	F_1	F_2		F_2 比例
圆型×皱缩子叶	圆形	5474圆	1850皱	2.96∶1
黄叶×绿色子叶	黄色	6022黄	2001绿	3.01∶1
紫花×白花	紫花	705紫	224白	3.15∶1
膨大×缢缩豆荚	鼓胀	882鼓	299瘪	2.95∶1
绿色×黄色豆荚	绿色	428绿	152黄	2.82∶1
花腋生×花顶生	腋生	651腋生	207顶生	3.14∶1
高植株×矮植株	高植株	787高	277矮	2.84∶1

第二节 孟德尔定律

孟德尔通过豌豆杂交对一对性状的观察得出了三条规律：①F_1 代的性状一致，通常和一个亲本相同。得以表现的性状为显性，未能表现的性状称隐性，此称 F_1 一致性法则。②杂种 F_1 代自交后产生的后代中，初始亲代的二种性状（显性和隐性）都能得到表达。③这两性状的比例总为 3∶1。

1900 年起，有六位遗传学家先后都重复孟德尔的杂交实验，结果几乎完全一致（表 1-2）。以后各国的遗传学家又以不同的动植物材料来进行验证，发现孟德尔的推论具有普遍性，因此人们就把孟德尔的假说归纳为**分离定律和自由组合定律**，从此成为遗传学最基本定律。

表 1-2 6 位学者重复孟德尔植物杂交实验的结果

实验者	亲代	F_2		F_2 比例
		黄	绿	
孟德尔，1865	黄色×绿色子叶	6022	2001	2.96∶1
Correns，1990	黄色×绿色子叶	1394	453	3.08∶1
Tschermak，1990	黄色×绿色子叶	3580	1190	3.01∶1
Hurst，1904	黄色×绿色子叶	1310	445	2.94∶1
Bateson，1905	黄色×绿色子叶	11903	3903	3.05∶1
Lock，1905	黄色×绿色子叶	1438	514	2.80∶1
Darbishire，1909	黄色×绿色子叶	109090	36186	3.01∶1
总数	黄色×绿色子叶	134707	44892	3.01∶1

一、分离律（law of segregation）

1. 分离律的实质

孟德尔对单个性状杂交的结果进行了深入的分析，提出一系列假设，归纳起来

可以这样阐述：控制性状的一对等位基因在产生配子时彼此分离，并独立地分配到不同的性细胞中。

等位基因（allele）这一名词是由贝特森（1902）提出的，当时的概念是指位于一对同源染色体上.位置相同，控制同一性状的一对基因。这一概念只适用于高等真核生物及经典遗传学，不能适合原核生物及分子遗传，所以现在的概念是指一个基因由突变而产生的多种形式之一。贝特森还创用了**纯合子**（homozygote）、**杂合子**（heterozygote）。一个或几个座位上的等位基因相同的二倍体或多倍体称纯合子，不同时称杂合子。

"基因"（gene）这个名词是由丹麦的科学家约翰森（Johannsen）于 1909 年提出的，取代了孟德尔的遗传因子。同年他又提出**基因型**（genotype）和**表型**（phenotype）这两个名词，前者指的是生物的内在遗传组成，后者指的是可观察到的个体外在性状，是特定的基因型在一定环境条件下的表现。

现在我们就用这些名词来阐明孟德尔的推测。表型为圆形的性状是显性的，其基因型是 RR，而相对表型为皱缩的性状是隐性的，基因型为 rr，他们都是纯合子，产生配子时，一对等位基因相互分离，雌雄配子各带有 R 和 r 基因，结合时，形成了杂合体 Rr。由于 R 控制显性性状，在杂合体 Rr 中可以得到表达，而隐性的基因在杂合体的遗传结构中虽然也存在，但得不到表达，所以 F_1 杂合体的表型是圆形的。F_1 杂合体的雌雄配子各含两种基因 R 和 r，概率皆为 1/2，雌雄配子随机组合的结果产生了 F_2，基因型为 RR、Rr 和 rr，比例为 1∶2∶1。由于显隐性的关系，表型只有圆型和皱缩两种，比例为 3∶1，其中只有 1/3 圆型籽粒是纯合体（RR），其余 2/3 为杂合体（Rr）（图 1－1）。这个假设很好地解释了杂交实验的结果，但是否完全正确，还要在以上假设的基础上，通过回交来加以验证。F_1 杂合体的表型为圆型，基因型为 Rr，当和隐性亲本（rr）回交时（图 1－1），杂合体可产生两种基因型不同的配子，R 和 r，而皱缩的隐性亲本只能产生一种配子 r，假设配子的结合是完全自由组合的话，那么会产生两种表型不同的回交后代圆型和皱缩，其比例应为 1∶1，这是在实验前的根据分配律推测的结果，经实验证实完全符合推测，使以上假设充分地得到了验证。

2. 分离定律的意义

分离定律也称为孟德尔第一定律，其意义有二：

(1)具有普遍性，不仅植物中广泛存在，在其他二倍体生物中都符合这一定律。如人类单基因遗传性状和遗传病约有 4344 种（1988 年），如虹膜的颜色、头发的颜色及形状（曲直），眼、口、鼻的形态，能否尝出苯硫脲（PTC）的苦味等等都是遗传的性状。像多指（图 1－2）、侏儒（先天性软骨发育不全）（图 1－3）、裂手裂足、亨廷顿舞蹈病（Huntington's chorea）等都是显性性状，有的性状可以传递很多代成为某些家族的特征，如 Hapsburg 王朝这一家族中的许多成员都有一窄而突出的下颚和使嘴成为半开的突出下唇。由于有画像存在，这个性状可一直追溯到 14 世纪。隐性性状常见的有白化、半乳糖血症、苯丙酮尿症、全色盲、早老症、自毁容貌综合征等。半乳糖血症是一种乳糖代谢的先天性缺陷，患儿缺乏半乳糖－1－磷酸尿苷酰转移酶（galactose－1－phosphate uridyl transferase，Gal－1－*PUT*），不能分解乳糖代谢过程中生成的半乳糖－1－磷酸。因此当患儿摄入含乳糖或半乳糖的人奶或牛

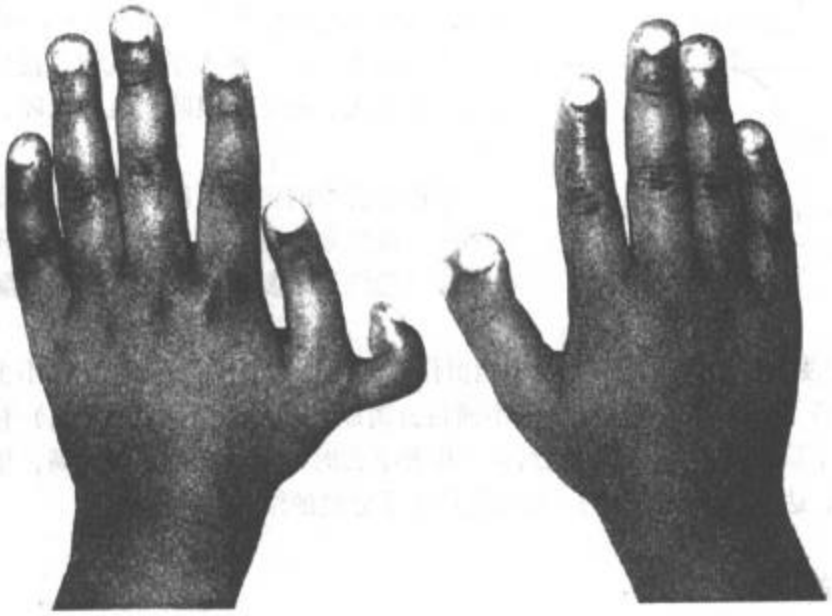


图 1-2 人类的显性遗传性状——多指。（转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5 th ed., 1993, Fig. 2-20）

图 1-3 以色列的一个具有先天性软骨发育不全（显性遗传性状）家族。（转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 2-17）

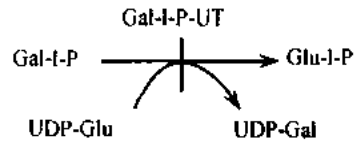


图 1－4　半乳糖血症的代谢障碍。
Gal:半乳糖；Glu:葡萄糖

奶后，血中半乳糖显著升高，乳糖－1－磷酸在细胞内聚集（图 1－4），抑制了糖代谢，使得智力发育迟缓，肝肿大、硬变，眼睛产生白内障，患儿往往夭亡。

动物中兽类的皮毛颜色，鸟类羽毛颜色的遗传也都符合孟德尔法则的，黑腹果蝇的长翅（显性）和残翅（隐性），复眼的颜色，身体的颜色等也都如此。

(2)使人们知道，杂合体是不能留作种子的，20 世纪 50 年代我国不少农民育种家，进行了大量的杂交实验，但不懂得分离原理，看到杂种后代（F_1）有很多优良品质，于是就留下做种子，但到第二年种下去时得到的结果是高的高，矮的矮，性状不同，达不到预期的目的，这就是产生了分离的结果。

3. 人类的家谱分析

家谱（pedigree）分析是研究人类遗传的重要方法之一，首先要能确定家谱中的亲缘关系，表示方法如图 1－5 所示。在家谱调查中第一次被发现的具有某种异常性状的人称**先证者**（proband or propositus）。表型正常而带有异常隐性基因的人称**携带**

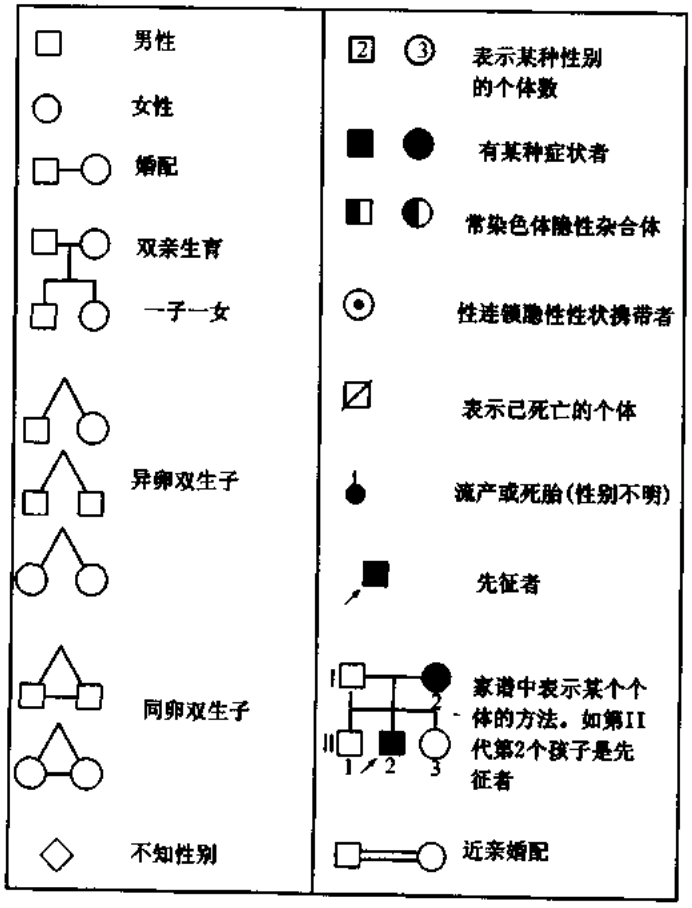


图 1－5　人类家谱分析中常用的符号。（转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 2－14）

者。我们从一个精确的家谱中能通过分析而推测出某一性状的遗传规律以及某些个体的基因型，而且还可以估计未来孩子可能患病的概率。

二、自由组合定律（law of independent assortment）

1. 自由组合律的内容

孟德尔在实验中不仅观察一对性状的遗传，而且还仔细观察分析了两对性状的遗传，也就是双杂种的杂交情况，具体情况如 图 1－6 所示。孟德尔将黄色圆型的豌豆和绿色皱缩的豌豆杂交，黄色对绿色为显性，圆型对皱缩为显性，杂交一代的表型全部为黄色圆型。F_1 代自交后产生了四种不同表型的籽粒（F_2）代，黄色圆型籽粒为 315 个，黄色皱缩的 101 粒，绿色圆型的 108 粒，绿色皱缩的 32 粒，它们之间的比近似于 9∶3∶3∶1。其中和亲本性状组合相同的后代，黄色圆型和绿色皱缩的称**亲组合**（parental combination），而不同于亲本形状组合的后代，黄色皱缩和绿色圆型的称**重组合**（recombination）。为什么会产生这样的比例呢？孟德尔进行了富有逻辑性的分析，首先只看其中的一对性状，如亲本为黄色绿色籽粒的遗传，F_1 代全为黄色，F_2 代 416 粒为黄色，140 粒为绿色。两者之比正好符合 3∶1。再看另一对性状圆型和皱缩，F_1 代全部呈现为显性性状圆型，F_2 代 423 粒为圆型，133 粒为皱缩，两者之比为 3∶18，也符合 3∶1。看来各种情况并不混合，而是独立遗传的。在 F_2 代中除了亲组合以外，又出现了重组合，而且各种表型籽粒的比为 9∶3∶3∶1，正是（3+1）的二次方展开式，孟德尔推测这是两对性状自由组合的结果。后人将其归纳为**孟德尔第二定律**：自由组合定律，即在配子形成时各对等位基因彼此分离后，独立自由地组合到配子中。

2. 自由组合的解释

与分离定律相同，黄色圆型亲本为显性性状，基因型为 RRYY；只能产生一种基因型的配子 RY，绿色皱缩的亲本为隐性性状，基因型为 rryy，也只能产生一种配子，基因型为 ry。杂交后 F_1 代是杂合体，两种配子结合，基因型为 RrYy。当 F_2 产生配子时等位基因 Rr 彼此分离，等位基因 Yy 也同样如此，彼此自由组合，形成四种不同基因型的配子而且数目相等：RY、Ry、rY 和 ry。四种不同配子又随机结合受精，在 F_1 上结四种不同的豆粒，它们的基因型应为 9 种，表型为 4 种，比例为 9∶3∶3∶1(图 1－6)。

3. 测交验证

以上假设完满地解释了两对杂交的结果，但孟德尔仍进行了测交来加以验证，治学的态度十分严谨，根据原假设 F_1 杂合子和显性亲本回交后代应全为黄色圆型；而和双隐性亲本测交，后代应为黄圆、黄皱、绿圆和绿皱四种基因型和表

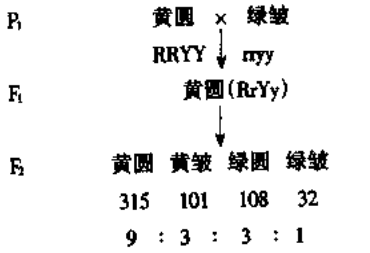


图 1－6 豌豆两对性状的杂交实验。

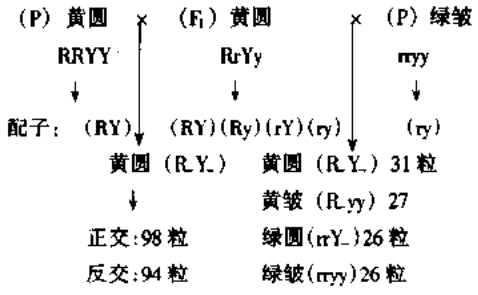


图 1－7　豌豆两对性状的回交实验。

型，其比例为 1∶1∶1∶1，而实验结果完全符合预期的结果，分别为 31 粒、27 粒、26 粒和 26 粒，接近于理论比（图 1－7）。

4. 多对基因的杂交

孟德尔不仅分析了两对基因的杂交，还研究了三对基因的杂交，从而推导出多对基因（n）"组合系列的项数（3^n）"，"系列的个体数 4^n"及保持稳定的组合数（2^n），如原种有 4 个性状不同，这个系列就具有 $3^4 = 81$ 个类别，$4^4 = 256$ 个体和 $2^4 = 16$ 个稳定的类型；也就是说每 256 个杂种后代中，有 81 种不同组合，其中16 种是稳定的。如果把这段论述用现代的术语来表达，那就是 3^n 个 F_2 的基因型，2^n 个 F_2 的表型，4^n 个 F_1 配子的组合数。根据这个推论可以再延伸一下，列成表 1－3。

表 1－3　杂合体自交或互交时所产生的各种基因型表型数以及分离比

基因对数	F_1配子类型	F_1配子组合	F_2基因型	F_2表型	分离比
1	2	4	3	2	$(3+1)^1$
2	4	16	9	4	$(3+1)^2$
3	8	64	27	8	$(3+1)^3$
4	16	256	81	16	$(3+1)^4$
N	2^n	4^n	3^n	2^n	$(3+1)^n$

5. 孟德尔自由组合定律的意义

（1）自由组定律也同样具有广泛的存在，比如有一个很有趣的昆虫行为遗传的例子，那是蜜蜂的卫生品系，此品系的蜜蜂可识别患了腐臭病（foul brood）的幼虫，将蜂房的盖子打开，拖出患病的幼蜂再扔掉，使整个蜂巢保持清洁。但还有另一种不卫生品系，不具有以上行为。将这两种蜜蜂杂交，产生的后代（F_1）全部是卫生品系，F_1 互交产生的 F_2 有四种类型，一种是卫生品系，另一种是只会揭开蜂房的盖子，而不会拖出病蜂，还有一种类型是不会揭开盖子，如果人们帮它将盖子打开，这些蜜蜂会将病蜂拖出，第四种类型是不卫生品系，既不会揭开盖子，也不会将病蜂拖出。这四种类型的比例为 9∶3∶3∶1，符合自由组合定律（图 1－8）。

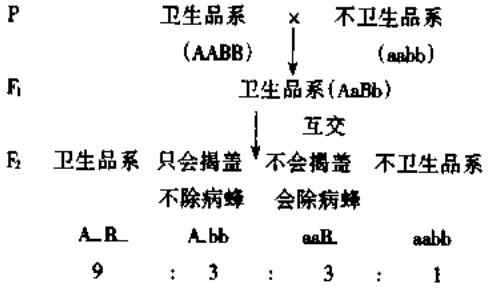


图 1－8　蜜蜂卫生品系行为的遗传。

(2) 由于自由组合的存在使各种生物群体中存在着多样性，使世界变得丰富多彩，使得生物得以生存和进化。

(3) 人们将自由组合的理论可以应用于育种，将那些具有不同优良性状的动物或植物，通过杂交使多种优良性状集中于杂种后代中，以满足人类的需求。

第三节　统计学原理在遗传学中的应用

一、概率的概念

概率（probability）又称**几率**（chance），是指某事件未发生前人们对此事件出现的可能性进行的一种估计。概率的公式为：

$$P_{(A)} = \lim_{n \to \infty} \frac{n_A}{n}$$

$P_{(A)}$：A 事件发生的概率。　n：群体中的个体数或测验次数。

n_A：A 事件在群体中出现的次数。

它和**频率**的概念不同，频率是指某一事件已发生的情况，如人口出生率的统计，升学率的统计，失业率的统计等等，这些事件都是已发生过的，数据都是经过真实调查得来的，不是估计的。但某事件以往发生的频率也可以作为对未来事件发生的可能性的估计。概率开始于赌博，所以开端的名声不好，三百多年前，一些赌徒问伽利略，一次掷三粒骰子，计算总点数，为什么出现总和为 10 的情况要比总和为 9 的情况多。1654 年一个赌徒问著名的数学家和哲学家巴斯加，把两粒骰子掷 24 次，出现两个 6 的情况会有多少。尽管概率论的“出身”不好，但现在对人类很有用，同样也是遗传学的一个重要理论基础。

概率有两个定则：①　$p + q = 1$，也就是说一件物若有两方面，那么它们发生的可能性总和等于 1。如足球裁判掷硬币，要么是正，要么是反，可能性各为 50%，总和为 1。②　$0 \leqslant p \leqslant 1$，当概率（$p$）= 1 时，称为**必然事件**，如“人终有一死”，“天总有晴日”等，$p = 0$ 时，称为**不可能事件**，如“太阳西出”，“永生不死”等。当 $0 < p < 1$ 时称可能事件，也是**随机事件**。有的事件可能性比较大，比如有人说“邻居的孩子很聪明，长大了能成为一名好学生”，这完全可能，如果有人说“这孩子长大了能当宰相”，这也是可能的，但概率很小很小。

二、概率规则

1. 相乘定律

独立事件是指两个或两个以上互不影响的事件，比如掷硬币，你第一次掷的结果可能是正，也可能是反，第二次掷也是如此，但其结果丝毫不受第一次结果的影响，是互相独立的事件。**独立事件** A 和 B 同时发生的概率等于它们分别出现的概率乘积。以公式表示如下：

$$P(A \cdot B) = P(A) \times P(B)$$

例如掷硬币，第一次出现正面的机会是 1/2，第二次掷出现正面的机会也是 1/2，现在要问两次都出现正面的机会是多少？那么就可用相乘定律：

$$P(A \cdot B) = P(A) \times P(B) = (1/2) \times (1/2) = 1/4$$

2. 相加定律

相加定律适合于互斥事件（matually exclusive events）。互斥事件是指不可能同时发生的事件。如掷硬币结果不是正就是反，不可能正反同时出现；人们生孩子，除异常情况外，不是生男孩就是生女孩；掷股子掷的结果只会出现正六面体中的一面点数。这些都是互斥事件。出现互斥事件 A 或 B 的概率等于它们各自出现概率的和。公式如下：

$$P(A 或 B) = P(A) + P(B)$$

如某人要生孩子了，生男孩或女孩的概率为：

$$P(男或女) = P(男) + P(女) = 1/2 + 1/2 = 1$$

如掷股子，只掷一次是 1 点或 6 点的概率为：

$$P(1 或 6) = P(1) + P(6) = 1/6 + 1/6 = 1/3$$

3. 组合事件（combining events）

如果两个或两个以上的事件，既有独立事件又有互斥事件，则称为**组合事件**。如掷两次硬币，问出现同样图案的机会是多少？这两个事件既是独立的，两次掷互不影响，又是互斥的，因同样的图案可能是正面，也可能是反面，只要求相同，这样就存在着同是正面或同是反面的问题。

$$P = 1/2 \times 1/2 + 1/2 \times 1/2 = 1/2$$

也就是说得到同是正面或同是反面的机会为 1/2。

4. 条件概率（conditional probability）

男性和女性中都会出现色盲，如只要求女性中的色盲概率，即既是色盲又是女性的概率，即为有条件的概率。

条件概率的公式：

$$P(B \mid A) = P(AB)/P(A)$$

现以白化病为例，这是一种常染色体隐性遗传病。如果有两个白化基因携带者结婚，那么他们生育的子女表型正常但又携带有害基因的概率是多少？根据遗传学的知识，我们可以列出下列的表格，表明表型和基因型的列联关系。根据表 1－4 可以得到。

表 1－4　事件 A、A′、B 和 B′频数表

	正常(A)	白化(A′)	合计
杂合(B)	$P(AB) = 1/2$	$P(A'B) = 1/4$	$P(B) = 1/2$
纯合(B′)	$P(AB') = 1/2$	$P(A'B') = 1/4$	$P(B') = 1/4$
合计	$P(A) = 3/4$	$P(A') = 1/4$	1

$$P(A) = 3/4;\ P(B) = 1/2;\ P(AB) = 1/2$$

那么表型正常的杂合体的概率即为在表型正常这个事件的条件下是杂合体的概率为：

$$P(A \mid B) = P(AB)/P(A) = (1/2)/(3/4) = 2/3$$

反之如要求杂合体中表型正常的概率为：

$$P(A \mid B) = P(AB)/P(A) = (1/2)/(1/2) = 1。$$

三、 概率的计算和应用

1. 棋盘法（punnett square）

棋盘法如图 1－9 所示，是一种最简单最直接计算杂交后代基因型和表型概率分布的方法。优点是准确可靠，缺点太烦琐，不适合多对基因的组合，如四对基因，配子类型有 32 种，将有 1024 种组合 81 种基因型，画起来太不方便，且容易出错，所以人们采用了比棋盘法要更为简便的分析法来推测。

♀配子 \ ♂配子	RY1/4	Ry1/4	ry1/4	rY1/4
RY 1/4	RRYY 1/16 黄圆	RRYy 1/16 黄圆	RrYy 1/16 黄圆	RrYY 1/16 黄圆
Ry 1/4	RRYy 1/16 黄圆	RRyy 1/16 绿圆	Rryy 1/16 绿圆	RrYy 1/16 黄圆
ry 1/4	RrYy 1/16 黄圆	Rryy 1/16 绿圆	rryy 1/16 绿皱	rrYy 1/16 黄皱
rY 1/4	RrYY 1/16 黄圆	RrYy 1/16 黄圆	rrYy 1/16 黄皱	rrYY 1/16 黄皱

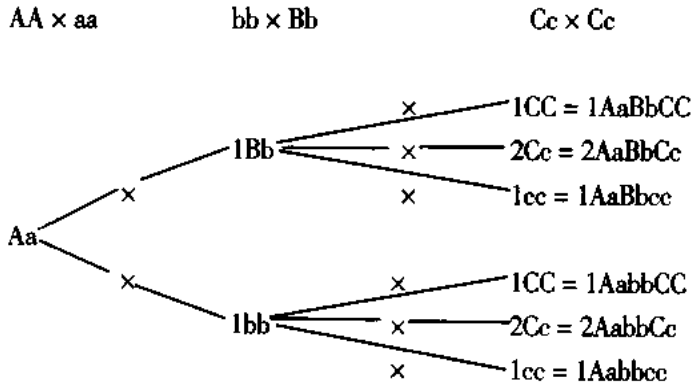
黄圆 9 ： 绿圆 3 ： 黄皱 3 ： 绿皱 1

图 1－9 豌豆两对性状分离比棋盘法图解（庞尼特方格）。

2. 分枝法（branching process）

如三对基因的杂交组合：AAbbCc × aaBbCc

我们可以将三对基因分别计算，再利用概率的原理推算三对基因杂交的结果：



3. 利用概率来计算多对基因杂交中某种基因型或表型出现的概率

若五对基因的杂交组合：

AABbccDDEe × AaBbCCddEe

求后代中基因型为 AABBCcDdee 和表现型为 A B C D e 的概率。

我们可以把每个基因作为独立事件来求五对基因同时发生某事件的概率：

P	AA × Aa	Bb × Bb	cc × CC	DD × dd	Ee × Ee
	↓	↓	↓	↓	↓
要求的基因型	AA	BB	Cc	Dd	ee
	↓	↓	↓	↓	↓
概率 p =	1/2 ×	1/4 × 1	×	1 × 1/4	= 1/32
要求的表型	A	B	C	D	e
	↓	↓	↓	↓	↓
概率 p =	1/2 ×	3/4 ×	1/2 ×	1/2 × 1/4	= 3/128

在计算时要注意基因型和表型的概率是不同的。

四、二项分布和二项展开法

我们研究一个小样品的一对质量性状时，如豌豆的紫花与白花，水稻的糯与非糯，人类的正常与白化，果蝇的长翅与残翅等等，各种情况的分布都呈离散型的不连续分布，这和数量性状变异的连续分布是不同的。这种离散型分布有两种类型：

1. 对称分布

如果一对性状各自发生的概率 p 和 q 相等时，分布是对称的。如生男孩的概率 $p = 1/2$，生女孩的概率 q 也为 1/2，$p = q = 1/2$

我们以有两个孩子的家庭为例，性别的分布可有以下几种情况：

第一个孩子	第二个孩子	概率	
男	男	1/2×1/2=1/4	P(pp)=1/4
男	女	1/2×1/2=1/4	2P(pq)=1/2
女	男	1/2×1/2=1/4	
女	女	1/2×1/2=1/4	P(qq)=1/4

$(p+q)^2=p^2+2pq+q^2=1/4+1/2+1/4$，分布是对称的。

若要研究三个子女的家庭，则为

$(p+q)^3=p^3+3p^2q+3pq^2+q^3=1/8+3/8+3/8+1/8$

分布也是对称的。

若研究 n 个子女的家庭，则为$(p+q)^n$。

2. 不对称分布

若一对性状各自发生的概率 $p\neq q$，那么二项式的分布是不对称的。

以隐性遗传病半乳糖血症为例，如果两个携带者婚配，只生两个子女，表型正常和患病的分布是：

第一个孩子	第二个孩子	概率	分布
正常	正常	3/4×3/4=9/16	P(pp)=9/16
正常	患儿	3/4×1/4=3/16	2P(pq)=6/16
患儿	正常	3/4×1/4=3/16	
患儿	患儿	1/4×1/4=1/16	P(qq)=1/16

3. 计算单项概率

若我们研究的不是两个孩子的家庭，而是多个孩子的家庭，也就是说二项式的方次数高时，往往我们要求的不是其全部，而是某一项的概率。

如白化病的携带者婚配，生育四个孩子的家庭中，白化与正常的分布情况每一单项的概率的计算可用二项分布概率密度函数公式：

$$\frac{n!}{x!(n-x)!}p^xq^n-x$$

$\frac{n!}{x!(n-x)!}$ 是组合公式，比如四个孩子中只有一个白化儿，他可能是老大，也可能是老二、老三或老四，就存在一个组合的问题，所以要加组合公式，p 为正常表型的概率，q 为白化发生的概率，x 为家庭中表型正常儿的数目，n 为家庭中孩子总数，本例中为 4，$(n-x)$则为家庭中患儿数，根据隐性遗传的规律我们可以得出表 1-5。

表 1-5　白化基因携带者婚、生育 4 个孩子的家庭中白化的频率分布

正常	白化	$\frac{n!}{x!(n-x)!}p^xq^n-x$	P
4	0	$1\times(3/4)^4(1/4)^0$	81/256
3	1	$4\times(3/4)^3(1/4)^1$	108/256
2	2	$6\times(3/4)^2(1/4)^2$	54/256
1	3	$4\times(3/4)^1(1/4)^3$	12/256
0	4	$1\times(3/4)^0(1/4)^4$	1/256

各项前面的系数也可用杨辉三角（图 1－10）来求出。

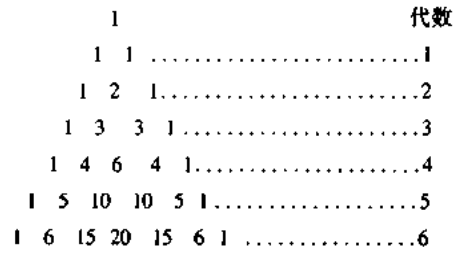


图 1－10　杨辉三角和二项分布概率密度公式系数的关系

以第二排为 1，一直向下排，如例子中为 4 个孩子，那么就第四排系数即为 1，4，6，4，1。要是六个孩子则为 1，6，15，20，15，6，1。

这一结果可以看出，白化的频率分布也是非对称的，也就是说，以以上婚配家庭，在 256 家中生了四个表型正常孩子的可能有 81 家，四个孩子中一个患白化病的可能为 108 家，两个患白化病的可能有 54 家，四个孩子都患白化病的可能有一家。

这个公式在遗传病的调查中是很有用的。因为像四个孩子都正常的这一项，是无法确定的，只能根据公式对这种情况进行估计，来修正我们的理论比。也就是说对某种有遗传倾向的特征和疾病的调查时，主要是统计父母表型正常，子女中有患者的各类家庭，正常儿和患儿之比是否符合孟德尔比(3：1)，如不加以校正的话，将会把第一类父母为杂合体，但子女中未出现患儿的家庭漏掉，比如就以上述白化病为例，我们只能根据是否生育患儿来判断父母是否为携带者，因此在有患儿的家庭中

正常儿为 $108\times3+54\times2+12\times1=444$
白化儿为 $108\times1+54\times2+12\times3=256$
正常儿：白化儿 $=1.734:1\neq3:1$

这是因为漏掉统计了 $81\times4=324$ 个正常儿。在实际调查中这一部分数据是无法调查的，因此可以用公式校正理论比，如上面的情况应是 1.734：1，而不是 3：1。

五、好适度的测验（goodness of fit）

1. 实际数和理论数的符合程度

我们做实验时所得到的数值不太会正好符合理论值，常常会出现一些偏差，比如在一个袋子里放入等量的黑白围棋子，充分混匀后让你闭着眼睛抓出 20 粒，按理论说应当是 10 粒黑子，10 粒白子，但你有可能抓出来的不是 10：10，而产生各种偏差，这些偏差应当是在一个允许范围内，那么什么是允许范围呢？什么样的偏差仍然说明原来放进的围棋子是 1：1 呢？

我们可以利用二项分布概率密度公式计算出二项式展开的任何一项的值，比如让我们在袋子里抓围棋子，一次只抓 2 粒，那么黑：白 = 1：1 的机会应是 0.500，

若一次抓 10 粒，那么正好 5 粒黑，5 粒白的机会是 0.246，一次抓 40 粒的话，符合 1∶1 的机会只有 0.125；一次抓 80 粒的话，符合 1∶1 的机会小到 0.089，依次下去越来越小，这种估计没有考虑存在偏差，叫“**点估计**”。点估计存在一些缺点：①单项的概率很小。比如让我一次抓 20 粒围棋子，不太会正好抓出是 10 粒白的，10 粒黑的，若抓出的是 14 粒白的，六粒黑的，按公式计算其得到的概率为 0.037，数字很小，凭借这个数据我们难以判断这个结果是否符合 1∶1。②样本越大，单项概率越小。在统计上样本越大结果应当越准确，但点估计却越小，似乎是矛盾的，所以要采用考虑偏差存在的“**段估计**”。此是计算从低到高的一个范围，而不是一项，即计算实得比数的概率和比实得比数偏差更大的比数概率的总和，如图 1－11 所示，如 14∶6 的点估计为 0.037，段估计为 $p = (0.037 + 0.015 + 0.005) \times 2 = 0.116$，也就是 100 次实验中可能有 11.6 次的机会发生这样的情况，是完全可能的，这一结果也是符合 1∶1 理论值的。

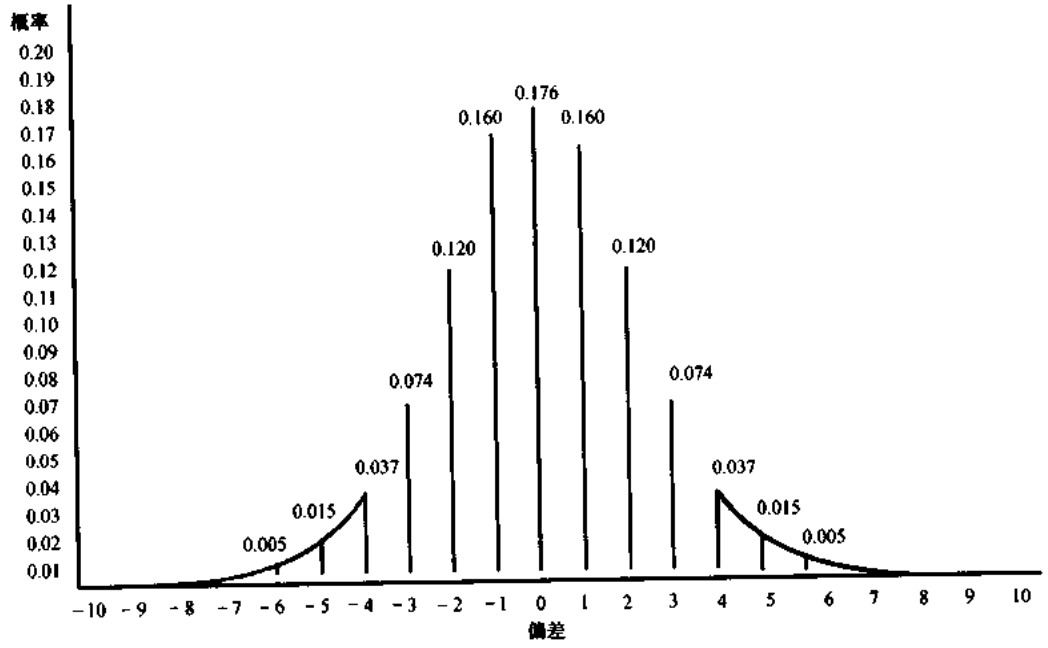


图 1－11　二项分布群体中点估计和段估计的关系。

（1）统计的标准

实验值和理论值的符合状况用什么标准来评价呢？统计学家规定了一个显著性标准：

$p > 0.05$ 结果与理论数**无显著差异**，实得值符合理论值。或实验组与对照组无差异。

$p < 0.05$ 结果与理论数**有显著差异**，实得值不符合理论值。或实验组与对照组有差异。

$p < 0.01$ 结果与理论数**有极显著差异**，实得值非常不符合理论值。或实验组与对照组有极显著差异。

（2）χ^2 测验

χ^2 （Chi-square method）也是一种好适度测验的方法，可以用来检验我们的实

验结果和理论值有无差异，其公式为：

$$\chi^2 = \sum(d^2/e)$$

e：为**预期值**，即**理论值**；d：预期值和**观察值**之差。例如我们做了以下两组豌豆测交（表 1－6）：查卡方表（表 1－7），一对性状，自由度 N＝2－1，$\chi^2 = 10$，$p < 0.01$ 有极显著差异；$\chi^2 = 2$，$0.05 < p < 0.30$ 无显著差异，符合理论比。

表 1－6　χ^2 值的计算〔紫花(Rr)×白花(rr)〕

	40		200	
	紫	白	紫	白
观察值	30	10	110	90
预期值	20	20	100	10
差数	10	10	10	10
d^2	100	100	100	100
d^2/e	5	5	1	1
$\chi^2 = \sum(d^2/e)$	10		2	

以上实验样本不同，虽然差数 d 相同，但结果明显不同，大样本的结果符合测交的理论值，小样本的结果却和理论值有极显著差异。因此在使用卡方测验时应注意以下几点：① χ^2 测验应用于大样本，当遇到小样本时，可用二项分布的方法来分析。② 预期数不得小于 5；③ 所取数值不用百分比表示。

表 1－7　χ^2 的数值表

df	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.00016	0.00098	0.0039	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.115	0.216	0.352	0.854	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.872	1.24	1.64	2.20	3.45	5.35	7.85	10.6	12.6	14.4	16.8	18.5
7	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.68	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2

第二章

遗传的染色体学说

基本概念

细胞是多细胞生物最小的“建筑”单位，细胞只可能由另一个细胞通过分裂而产生。

原核生物是最简单的单细胞生物，如细菌。原核生物缺乏由膜包被的核，以二等分裂（binary fission）来繁殖。原核的遗传物质是单个的环状的 DNA 分子，DNA 上结合有少量的蛋白质。

真核生物的细胞中，遗传物质位于有膜包被的核中，遗传物质分布在各条线状染色体上，每条染色体都由很多蛋白质和 DNA 组成。

二倍体细胞具有二套染色体，分别来自于两个亲本，每对染色体的两个成员被称为同源染色体（homologous chromosome），单倍体真核细胞仅有一套染色体。

有丝分裂是二倍体或单倍体真核细胞分裂的过程。有丝分裂产生两个子核，它们含有相同的染色体数和遗传信息。

减数分裂发生在有性生殖的真核细胞中。它是由二倍体细胞或核通过一次 DNA 复制和二次分裂产生 4 个单倍体细胞或核的分裂过程。

减数分裂的 4 个子细胞染色体数比母细胞减少一半。由于父本和母本的染色体的重新组合及染色体的交换，使 4 个子细胞所含的遗传信息有一定的差异。

遗传的染色体学说指出染色体是基因的载体。

第一节　有丝分裂和减数分裂

孟德尔的发现已经使我们知道生物的性状是由“基因”控制的，但这个基因是一种抽象的因子，接着人们就必须了解它是否存在？如果存在的话位于细胞中哪一个部位？它的结构如何？19 世纪末，Flemming W（1882）和 Boveri T（1891）分别发现了**有丝分裂**（mitosis）和**减数分裂**（meiosis），为遗传的染色体学说提供了理论基础，遗传的染色体学说现在看来似乎比较简单，但在遗传学的发展上却是十分重要的一步，这种推理的准确性也堪为楷模。

染色体学说是怎样形成的呢？证据逐渐积累来源多种多样。首批证据之一是来自细胞核分裂时染色体行为的发现。细胞的分裂有两种不同的类型：有丝分裂和减数分裂。在孟德尔研究和孟德尔定律的再发现之间（1886～1900）有很多的生物学家并不知道孟德尔的发现，但对遗传的现象很感兴趣，而且用完全不同的方法来研究这个问题。这些研究者想知道遗传物质在细胞的什么部分。很显然值得寻找的地方是配子，因为它们是亲子之间惟一的连联，卵和精子的大小虽不同，但人们相信它们对于后代遗传的贡献是相同等的。卵细胞质的量很多，而精子很小，其细胞质不像是遗传物质的所在地。然而卵和精子的核大小相同，因此被认为是遗传结构较合适的藏身之地。

细胞核的主要成分是什么呢？很明显是染色体，它在所有的细胞中都存在。这

一特点强烈地吸引着生物学家们，他们最为关注的是在一个生物体中不同细胞之间染色体数的恒定性。那么染色体数目的恒定性是怎样维持的呢？要想回答这个问题只有在有丝分裂和减数分裂时用显微镜来观察染色体的行为。从观察结果中发展成遗传的染色体学说——染色体上带有基因。

一、有丝分裂

细胞周期（cell cycle）是从一个细胞的任何一个分期阶段到子细胞的同等阶段中的各种事件。

从一个新产生的细胞到它分裂产生子细胞这一过程称为细胞周期，它可以分成为四个阶段：**M 期、S 期、G_1 期和 G_2 期**。M 是分裂期，通常是细胞周期中最短的时期，约占整个时期的 5～10%的时间。DNA 的合成发生在 S 期（synthesis）。G_1（gap_1）和 G_2（gap_2）是 S 期和 M 期之间的两个间隙期。G_1、S、G_2 合称为**间期**（interphase），即两个分裂期之间的时期。此期染色质均匀地分布于核中，所以在显微镜下看不到染色体。

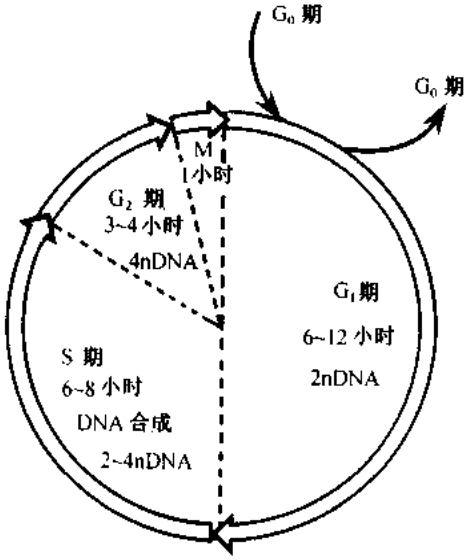


图 2－1　细胞周期。细胞周期中的间期由 G_1、S 和 G_2 三个阶段组成。分裂期在两个细胞周期之间。有时细胞可从 G_0 期进入细胞周期，也可在分裂后进入 G_0 期。（参考 B.Lewin：Genes Ⅵ，1997，Fig.36－1）

染色质网形成时核中的染色体已完成了复制，然后再纵裂分开，形成两条染色体，每一条进入不同的子核中，有丝分裂产生了两个完全相同的子核，分裂成两个**子细胞**。为了便于研究科学家们把这一连续的过程划分为四个阶段，称为**前期**(prophase)、**中期**(metaphase)、**后期**(anaphase)和**末期**(telophase)。

1. 前期

此期染色体开始逐渐变得清晰可辨，逐渐凝缩使其缩短变粗，收缩成螺旋状，这种形状易于移动。当染色体变得明显可见时，每条染色体已含有两条**染色单体**(chromatids)（图 2－2c），互称为**姐妹染色单体**(sister chromatids)。通过**着丝粒**(centromere)把他们相互连接在一起，到前期末，**核仁**(nucleoli)逐渐消失，核膜开始破裂，核质和细胞质融为一体。

2. 中期

在此期**纺锤体**（spindle）逐渐明显，这个鸟笼状的结构在核区形成，由细胞两极间一束平行的纤丝构成。着丝粒附着在纺锤丝上，染色体向细胞的**赤道板**

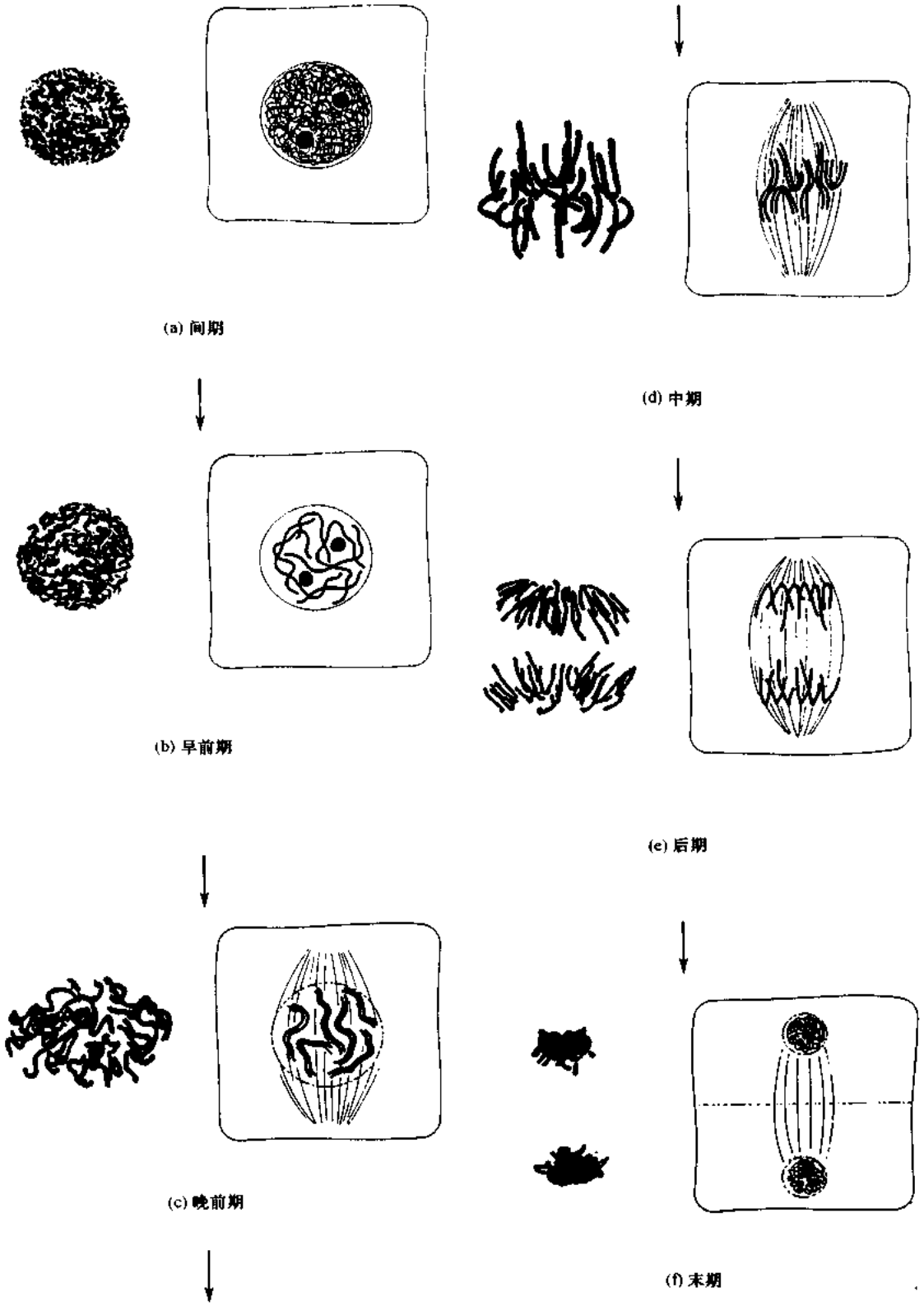
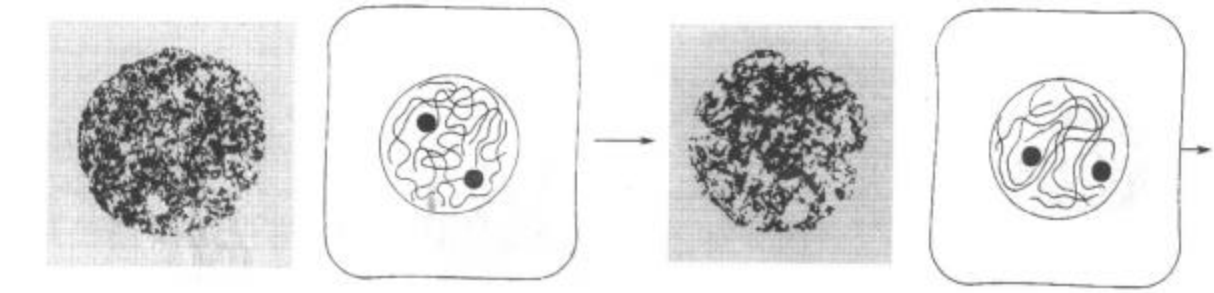
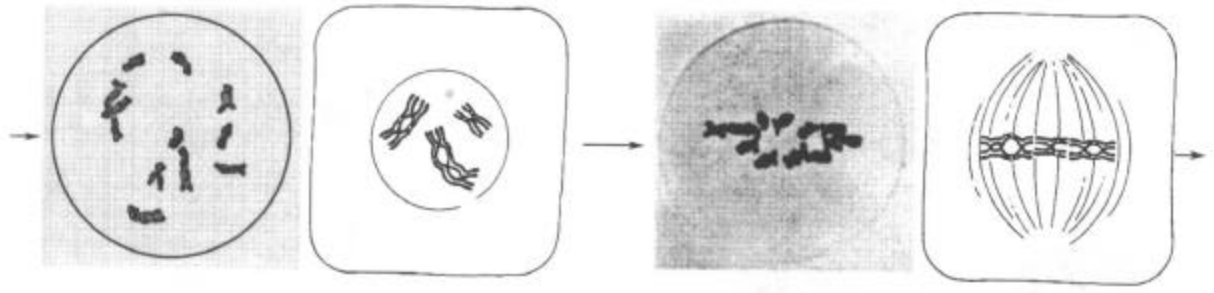


图 2-2　百合(*Lilium regale*)根尖细胞有丝分裂的图像。

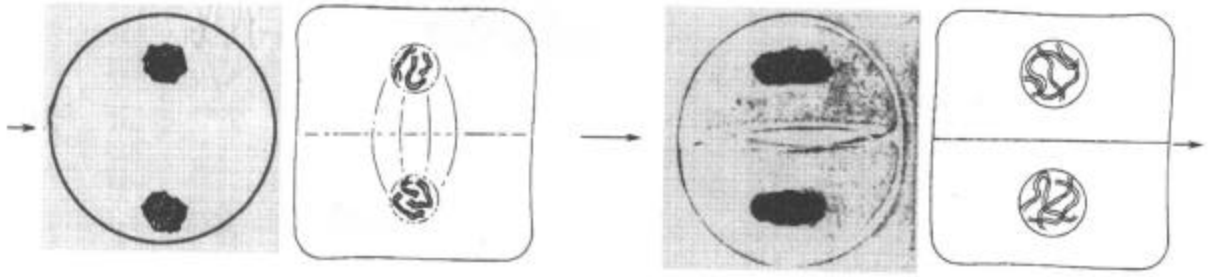
(转引自 Griffiths A.J.F. *et. al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-2)



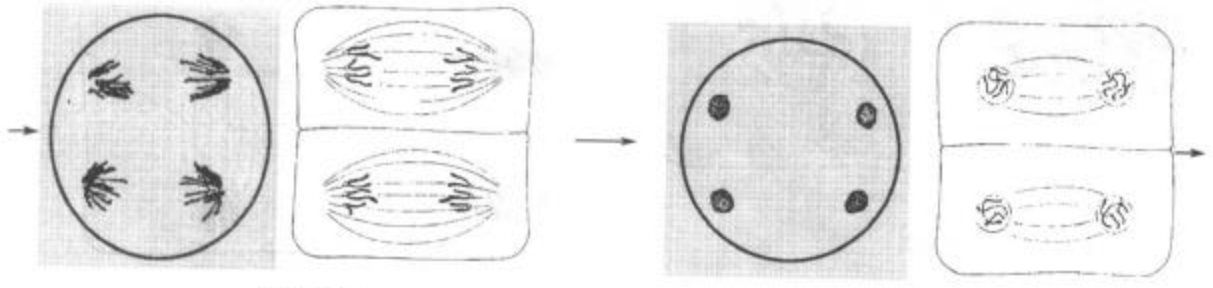
(a) 细线期　　(b) 偶线期



(c) 终变期　　(f) 中期 I



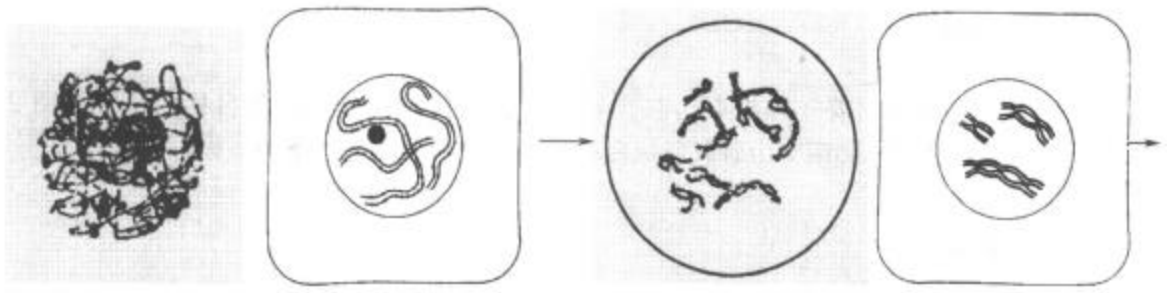
(i) 末期 I　　(j) 间期



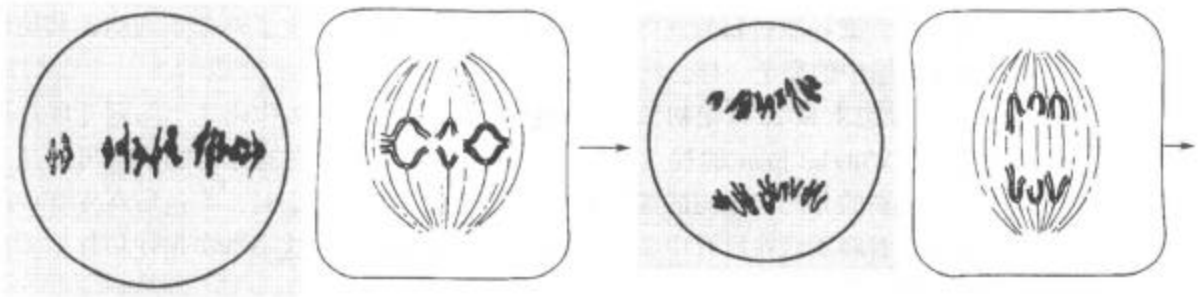
(m) 后期 II　　(n) 末期 II

图 2-3　减数分裂和花粉的形成

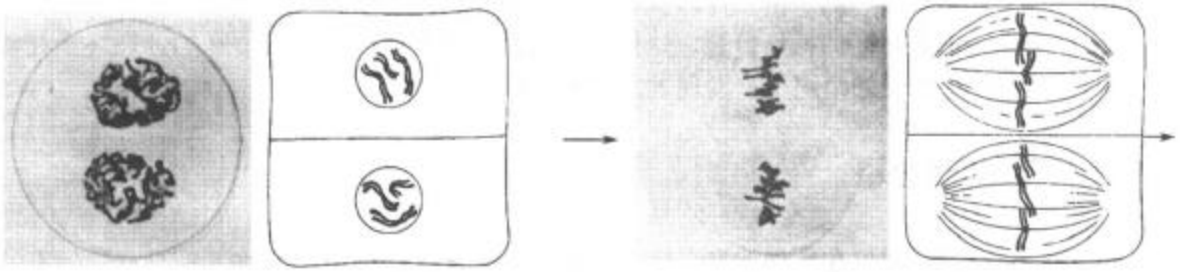
(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-3)



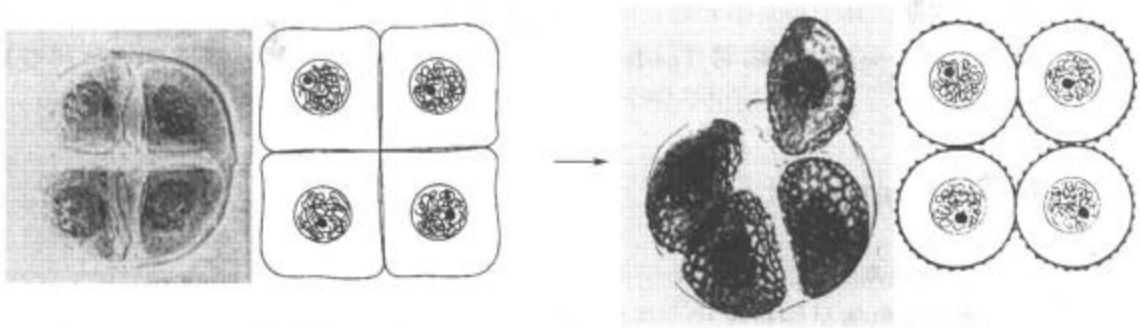
(c) 粗线期　　(d) 双线期



(g) 早后期 I　　(h) 晚后期 I



(k) 前期 II　　(l) 中期 II



(o) 四分体　　(p) 花粉粒

（equatorial plane）移动（图 2-2d）。

3. 后期

在后期，着丝粒纵裂为二，姐妹染色单体彼此分离，各自移向一极（图 2-2e）。染色体的两臂由着丝粒拖曳移动，这时染色体是单条的，称为**子染色体**。

4. 末期

末期子细胞的染色体凝缩为一个新核，在核的四周核膜重新形成，染色体又变为均匀的染色质，核仁又重新出现，又形成了间期核（图 2-2f）。末期结束时，纺锤体被降解，细胞质被新的细胞膜分隔成两部分，结果产生了两个子细胞，其染色体和原来细胞中的完全一样。

在 19 世纪末和 20 世纪初，关于染色质在发育中的直接作用这个问题开展过激烈的争论，Wilhelm Roux 的观点认为，如果染色体在发育中起作用的话，可能是由于他们有区别的分配到不同的细胞系中。当这一观点被否定时，染色质在发育中的重要性也被看轻了。在发育中染色体物质完全保持恒定的这一事实是在核移植实验成功后才提出的。

有丝分裂维持了每个核中染色体数目的恒定被证实了，但这使得早期学者对于两个配子参与受精感到大惑不解。他们认为在受精时两个核融合会使后代的染色体加倍，但在每一代中染色体的数目却永远保持恒定，感到无法解释。有人推测存在一种染色体减半的特殊分裂，这样可以解释上述的问题。这种特殊的分裂终于在动植物产生配子的组织中被发现了，称为减数分裂。

二、减数分裂

减数分裂仅在性母细胞进行。减数分裂中大部分 DNA 都是在 S 期合成，但有的是在减数分裂的前期合成的。减数分裂含有两次分裂，称为**第一次减数分裂**和**第二次减数分裂**。两次减数分裂的特点是不同的。每次减数分裂都可以分成前、中、后、末四期，其中最复杂和最长的时期是前期Ⅰ，又可分为**细线期**（leptotene）、**偶线期**（zygotene）、**粗线期**（pachytene）、**双线期**（diplotene）和**浓缩期**或称**终变期**（diakinesis）。这些时期也是完全连续的过程。

1. 前期Ⅰ（prophase Ⅰ）

(1)细线期

此期染色体呈细长线状，核仁依然存在（图 2-3a）。在细线期和整个的前期中染色体持续地浓缩。细线期中，沿着每条染色体浓缩的小区域称为**染色粒**（chromomeres），呈链珠状。

(2)偶线期

细线状的染色体开始配对，在性母细胞中，实际上有两套染色体，每条染色体都有一条与之**同源**（homologous）的染色体，逐步相互配对或**联会**（synapsed），就

像“拉链”式的两个成员称为**同源染色体**（homologs），在有丝分裂中是没有这一过程的。在有丝分裂中无论细胞中有几组（X）染色体，在减数分裂中只分成两套（即两套基因组）。带两套染色体的细胞就称为**二倍体**（diploid）或“**2n**”表示，即是染色体的套数。只带有一套染色体的细胞就称为**单倍体**（haploid）。在大部分高等生物中如人类和显花植物的细胞都是2倍体。性母细胞经减数分裂所形成的配子只有一套染色体，是单倍体细胞。

在偶线期一对同源染色体是怎样互相识别的呢？可能由于染色体末端的端粒（telomeres）是被锚定在核膜上，而同源染色体的端粒彼此相靠得很近，拉起“拉链”时可能从端粒处开始。“拉链”又是怎样拉的呢？两条同源染色体精确配对的机制是怎样的呢？虽然这个机制目前还未完全弄清，但我们已知需要一种由蛋白质和DNA所构成的精巧结构，称为**联会复合体**（synaptonemal complex，SC）（图2-3b），联会时它总是夹在两条同源染色体之间。

1973年Counce和Mayer建立了界面铺展技术使SC的研究得到了发展。人们建立大量的SC**组型**（karyotype）（图2-4），并试图揭示SC的形成、同源染色体配对及重组的分子机制，并且对SC蛋白及其编码基因进行了筛选、鉴定和特征分析。SC组型分析可用于系统分类、染色体畸变及致癌、致突变和致畸物的筛选等，有着广阔的应用前途。

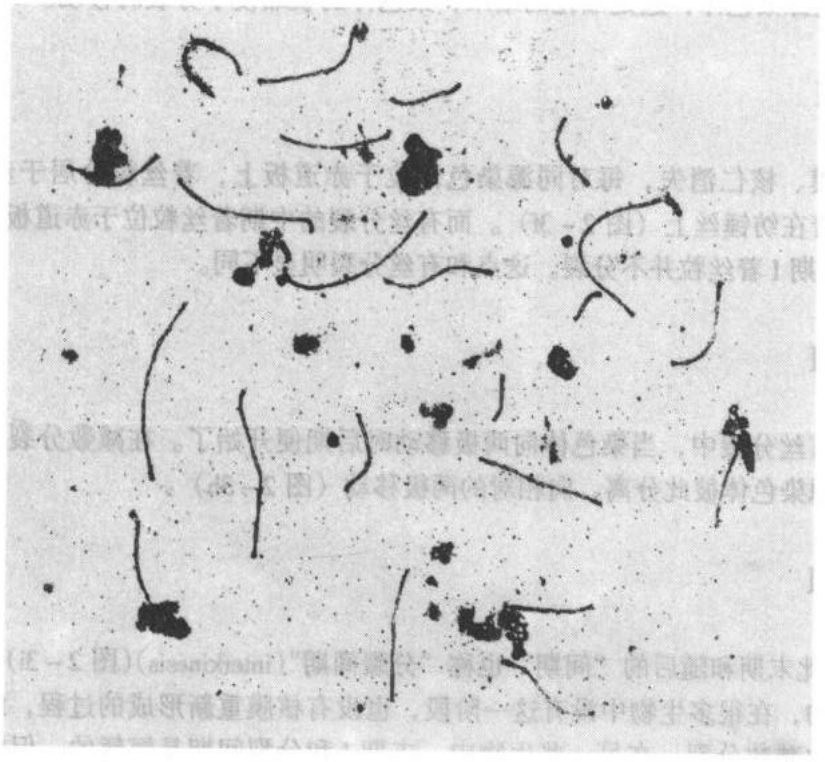


图2-4 用微铺展技术制备的SC核型。金黄地鼠（*Mesocricetus auratus*）的SC核型(照片由史庆华博士惠赠)

(3)粗线期

在这一阶段，染色体完全联会，缩短变粗，但核仁仍然存在。在核中同源染色体的对数等于n。一对配对的同源染色体称**二价体**（bivalent）或**四联体**（tetrad），在特殊情况，存在不能配对的染色体则称**单价体**（univalent）或**二联体**（dyad）。染色粒的直线排列在每一对的同源染色体上，像一串精致的珠链。

(4)双线期

在细线期时每条同源染色体看起来都是单条的线状，其实此时 DNA 已在 S 期复制过了，只不过难以分辨出而已。显然到了双线期每条染色体就出现两条（图 2-3d），与有丝分裂中期相似，其中每一条就称为**染色单体**（chromatids）。由于配对的同源染色体每条都产生两条姊妹染色单体，所以联会复合体的结构是含有一束四条染色单体，在双线期，同源配对时不太紧，其实是互相排斥呈现轻微的分离。非姊妹染色单体间的交错结构称为**交叉**（chiasmata 或 singular, chiasma），每一对同源染色体都有一个或更多的交叉存在。发生在减数分裂早期的交叉称为**交换**（crossovers），可以发生在偶线期或粗线期。交换主要存在于减数分裂，在有丝分裂中十分罕见。一个交换是两个非姊妹染色单体之间的一次精确的断裂、**互换**（swapping）和重接。人们研究了生物的异常品系，其交换很少或完全不发生，表明在减数分裂中染色体有序地分配入子细胞的过程被打乱了。这样交换很显然帮助决定配对染色体的行为，每对同源染色体至少发生一次交换，这对于正常的分离是必要的。交换另外的作用是产生新的基因组合，这是群体遗传变异的一个重要来源。交叉的端化，非姊妹染色单体之间要彼此滑动，这样延缓了染色单体的分离。

(5)终变期或称浓缩期

此期明显不同于双线期，染色体进一步地收缩，常可见到“O”形或“+”形的一对同源染色体，这是端化的结果，染色体的收缩便于分裂时移动。

2. 中期 I

核膜、核仁消失，每对同源染色体位于赤道板上，着丝粒分居于赤道板的两侧，附着在纺锤丝上（图 2-3f）。而有丝分裂的中期着丝粒位于赤道板上（图 2-3f），中期 I 着丝粒并不分裂。这点和有丝分裂明显不同。

3. 后期 I

在有丝分裂中，当染色体向两极移动时后期便开始了。在减数分裂的后期 I 是一对同源染色体彼此分离，向相对的两极移动（图 2-3h）。

4. 末期 I

此末期和随后的“间期”也称“**分裂间期**”(interkinesis)(图 2-3i)，并不是普遍存在的，在很多生物中没有这一阶段，也没有核膜重新形成的过程，细胞直接进入第二次减数分裂。在另一些生物中，末期 I 和分裂间期是短暂的，但核膜重新形成。在很多情况下，此期不合成 DNA，染色体的形状也不发生改变。

按照惯例，染色体到达两极，解旋后成为染色质，重新形成两个子核，每一个核是一个单倍体，这是因为它只含一套染色体，但每条染色体都含有两条染色单体附着在着丝粒上。也就是说我们可以以着丝点来计算染色体的条数，而不管是否有姊妹染色单体的存在。在每个子细胞中染色体减半这是个关键，因此减数分裂 I 被称为“**减数分裂**”（reduction division），而第二次减数分裂和有丝分裂相同，染色体的数目保持不变，这种类型的分裂称为“**等数分裂**”（equational division）。

5. 前期Ⅱ

此期和前期Ⅰ的情况相似，每条染色体都已经复制过，所不同的是此期的染色体数是 n （图 2－3k）。

6. 中期Ⅱ

染色体排列在赤道板上（图 2－3l），纺锤丝附着在单个的着丝粒上。染色单体从彼此紧密相联逐渐部分地分离。

7. 后期Ⅱ

着丝粒纵裂，姊妹染色单体由纺锤丝拉着向两极移动（图 2－3m）。

8. 末期Ⅱ

在子细胞两极染色体周围核膜重新形成（图 2－3n）。

减数分裂产生四个子细胞(图 2－3o),每个子细胞发育成花粉粒(图 2－3p)。在另一些生物中，四个子细胞分化为另一种结构，如动物的精细胞。

三、减数分裂的特点

①具有一定的时空性，也就是说它仅在一定的发育阶段，在生殖细胞中进行。而有丝分裂几乎在生物的一生中，在各种不同的组织中都进行，除少数的情况是例外的。如哺乳动物的脑细胞和肝细胞都不进行有丝分裂，肝细胞在损伤后可以进行分裂。②减数分裂经第一次分裂后染色体就进行减半，所以减数分裂的产物是单倍体。③前期长而复杂，同源染色体经历了配对、联会、交换，从而使遗传物质进行了重组。④每个子细胞遗传信息的组合是不同的。

四、染色体的类型

染色体的一般结构在细胞生物学中都有描述，在此不再叙述，而着重介绍染色体的各种类型。

按照染色体的形态，正常的染色体常根据长臂和短臂之比以及有无随体分为 5 种类型（表 2－1）。这在进行动植物核型（karyotype）分析时用来描述不同染色体的形态。核型是指某一物种染色体的组成，通常用中期染色体的照片，按长臂的大小或总的长度依次排列，用来表明物种的特点以及和亲缘种之间的进化关系，在系统分类中有着重要的参考价值。如在探讨熊猫的起源时，根据解剖结构形态特征，有人认为来源于浣熊，也有人认为来源于棕熊，各执一词，相持不下，结果经核型分析表明熊猫和棕熊的亲缘关系更近于浣熊，澄清了这一争论（参见第 11 章）。

表 2－1　核型分析中采用的染色体类型，染色体不同臂比的各种类型

名　　称	符　　号	臂　　比
中部着丝粒染色体(metacentric chromosome)	M.m	1～1.7
亚中着丝粒染色体(submetacentric chromosome)	S.m	1.7～3.0
端部着丝粒染色体(telocentric chromosome)	t	7.0～∞
亚端着丝粒染色体(subtelocentric chromosome)	st	3.0～7.0
随体染色体(satellite chromosome)	Sat	

按照染色体的大小，染色体可以分为**大型染色体**（macrochromosome）如人类的染色体和蚕豆（图 2－5）等显花植物染色体，长度可达几个 μm，另一种称**小型染色体**（microchromosome），如鸟类和某些植物（如石斛属）（图 2－6)的染色体，常呈点状，长度只有零点几个微米，难以度量。

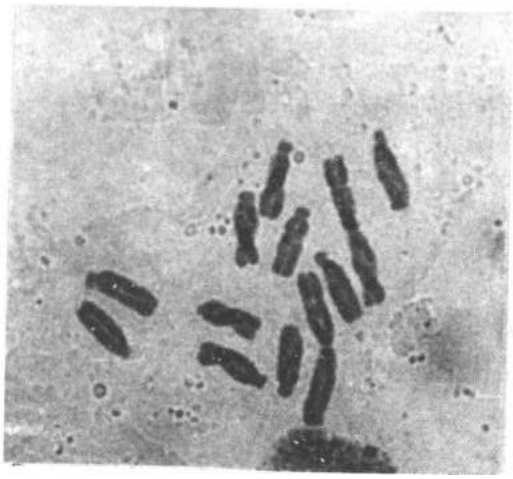


图 2－5　蚕豆(*Vicia faba*)的中期染色体(2n = 12)。

某物种的染色体数目分为**正常染色体**（A chromosome，**A 染色体**）和**超数染色体**（supernumerary chromosome 或 **B 染色体**）。在正常核型范围内的都称为 A－染色体。在细胞里出现多于正常核型以外的几条染色体，称 B 染色体，一般比 A 染色体小，虽有着丝点，但数目不稳

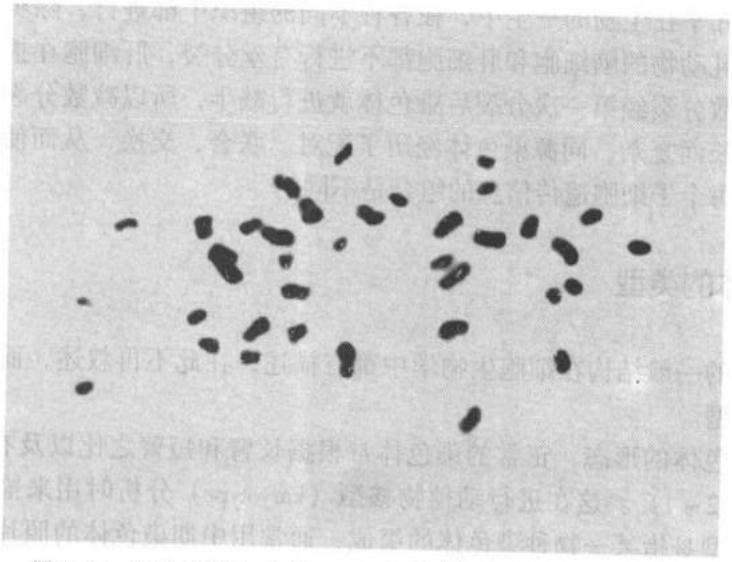


图 2－6　铜皮石斛(*Dendrobium crispulum*）的中期染色体（2n = 38）。

定，在有丝分裂中常常不是均等分配到子细胞中，其表型的效应也不明显。在很多种植物(如玉米、贝母等)(图 2－7)中以及少数种类的动物(如啮齿动物)中都观察到有 B 染色体的存在。

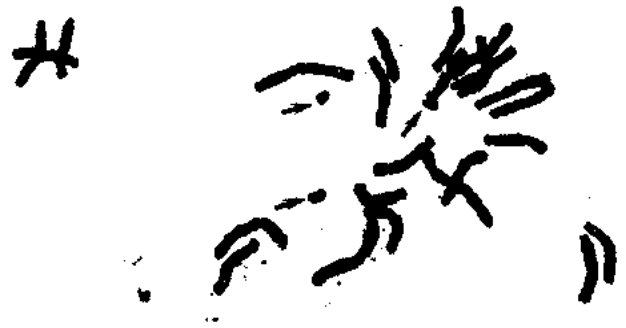


图 2-7 湖北贝母(*Fritilaria hupehensis*)中的超数染色体(BC)。(引自徐晋麟等:三种贝母核型的比较研究,植物分类学报,1987,25(1):77-80)

根据染色体的功能可以分成**常染色体**（autosome）和**性染色体**（sex chromosome）。常染色体上一般不具有决定性别的主要基因，一对同源的常染色体形态总是相同的；性染色体上具有决定性别的相关基因。在高等脊椎动物和某些植物中，一对性染色体是不完全同源的，形态大小也不同。

染色质也可分为**常染色质**(euchromatin) 和**异染色质** (heterochromatin) 区别如表 (2-2)。异染色质又分为**组成型**和**兼性**两类：组成型异染色质（constitutive hetero chromatin）无论在什么情况下都不能转录、表达，如着丝粒；而兼性异染色质（facultative heterochromatin）不同，在一定的时期是有功能的，如 X 染色体，在雌性哺乳动物发育早期（卵裂 5000～6000 细胞以前），两条 X 染色体都有功能，以后有一条 X 染色体失活，呈异染色体状态，也就是**巴尔小体**（Barr body）。

表 2-2 两种染色质的区别

常染色质	异染色质
间期染色淡	染色深
中期染色深	染色淡
染色体大部分区域	着丝粒附近
含基因	不含基因
复制早,可转录	复制迟,不转录
收缩程度大	收缩程度小

第二节 在生活周史中的减数分裂

一、动物的减数分裂

绝大多数多细胞动物在其大部分生活史中都是以二倍体存在的。在每种动物

中，经过减数分裂产生单倍体配子。通过受精，不同配子的核进行融合再形成二倍体的合子。经过不断地有丝分裂发育成一个新的二倍体生物。这样配子仅是生命周期中的单倍体阶段，配子仅在特殊的组织中发现。雄性动物产生的配子是**精子**。这个过程称**精子发生**（spermatogenesis）。雌性的配子是**卵**，由**卵子发生**（oogenesis）而形成（图 2－8）。

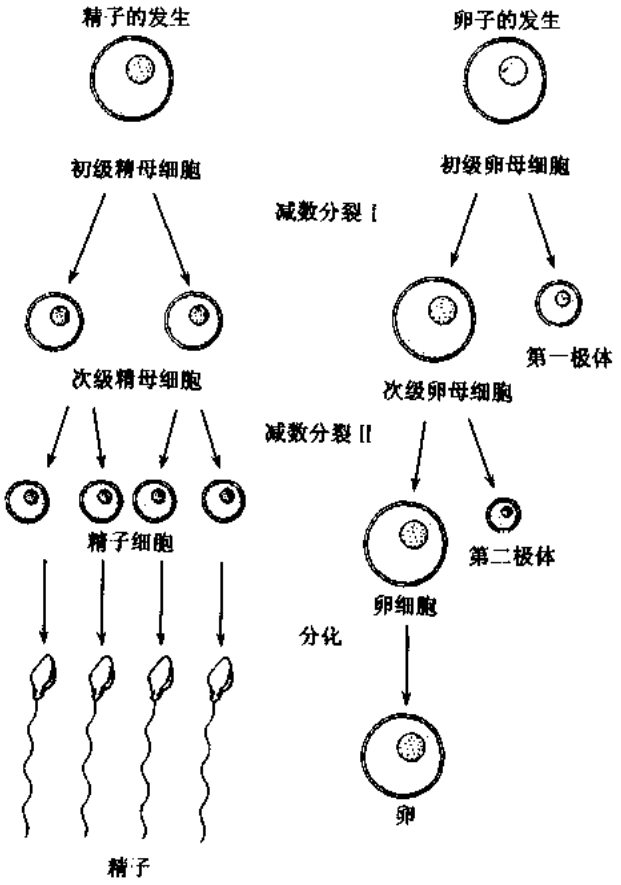


图 2－8　动物精子和卵细胞的形成。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 1－30）

在雄性动物中精细胞是由睾丸产生的。睾丸中含有**原生殖细胞**（primardial germ cell），也叫**初级精原细胞**，通过有丝分裂，产生**次级精原细胞**（secondary spermatogonia），由它转变成**初级精母细胞**（primary spermatocytes or meiocytes），每个精母细胞经过第一次减数分裂成两个**次级精母细胞**（secondary spermatocytes），经过第二次减数分裂每个次级精母细胞再分裂成两个**精细胞**（spermatids），这样一共产生 4 个单倍体精细胞。经进一步分化成成熟的雄性配子，即精子（spermatozoa）。

在雌性动物中卵巢含有**初级卵原细胞**（primary oogonia），通过有丝分裂产生**次级卵原细胞**（secondary oogonia），这些细胞再转化成**初级卵母细胞**（oocytes）。初级卵母细胞通过第一次减数分裂和不均等的**胞质分裂**（cytokinesis）产生两个细胞，大的一个称为**次级卵母细胞**（secondary oocyte），很小的一个称为**第一极体**（first polar body）。经过第二次减数分裂，次级卵母细胞产生两个单倍体细胞，一个是很小的细胞称为**第二极体**（second polar body）；另一个是较大的细胞，它迅速地发育为成熟的**卵细胞**（eggcell or ovom）。第一极体可分裂可不分裂。只有卵子是有生命力的

配子。这样在雌性动物中只能产生一个成熟的单倍体的卵（图 2-8）。

二、植物的减数分裂

在具有有性生殖生活周史的植物中有两种典型的时期。一种是**配子体**(gemetophype)时期或单倍体阶段，在此阶段产生配子；另一种是**孢子体**（sporophyte）时期或二倍体阶段，此期产生两倍体**孢子母细胞**。

在被子植物中，显花植物的花是有性生殖器官，通常它含有雌、雄两种生殖器：**雄蕊**（stamens）和**雌蕊**（pistils）。雄蕊含有单个的柄，花丝在柄的顶端，也就是花药。花药里有很多花粉，这是重要的雄性配子体。雌蕊含有雌配子体及典型的柱头，表面粘着，易于接受花粉。

减数分裂的一种特殊的方式在花的雄蕊和雌蕊中进行。在花的子房中，每一个胚珠是一个大的 2n 细胞，称为**大孢子母细胞**（megaspore mother cell），这些细胞经减数分裂产生两个单倍体产物，其中只有一个保持了生命力，这就是**大孢子**，或**单倍体雌性配子体**。大孢子的产生叫做**大孢子发生**（megasporogenesis）。

减数分裂在大孢子母细胞中通常涉及两次分裂，产生 4 个单倍体细胞。其中 3 个退化，只有一个大细胞能生存下来，它具有一个单倍体核，再经过 3 次有丝分裂，大孢子体带有 8 个相同的单倍体核，这 8 个单倍体核在细胞的两端，每端 4 个。然后每端有一个核回到细胞的中间部位。接着细胞壁产生，导致胞质的不均等分裂，中间的两个核的周围形成了大的双核细胞，受精时和一个雄核结合产生 3 倍体的**胚乳**（endosperm），因此称它为**胚乳母细胞**（endosperm mother cell）或叫做**极核**(polar nuclei)。其他的 6 个核各自被封闭在细胞特定位置上。整个具有八个核的位于靠近胚乳的端部的大配子体称为**胚囊**（embryo sac）。6 个单倍体细胞中的一个称**卵细胞**，它将成为胚珠中单个的雌性配子，它随机和一个精细胞结合而受精。

产生花粉粒或雄性配子体也要经过减数分裂，但要简单一些，这个过程称**小孢子发生**（microsporogenesis）。花药含有 4 个花粉囊，其中有很多**小孢子母细胞**（microspore mother cell）。每个小孢子母细胞经减数分裂产生一个单倍体的生殖核和一个单倍体的**管核**，然后产生一个硬壳而成为**花粉粒**。

受精时花粉粒附着在花柱头上，萌发后产生花粉管。花粉管生长穿过柱头进入花柱，在花粉管生长前产生的酶使组织降解，当花粉管伸入花柱，核停留在生长的顶端：单倍体的生殖核再经过一次有丝分裂产生二个精核或雄核。此时管核不再分裂。花粉管最后穿入胚珠上的一个小孔叫**珠孔**（micropyle），两个精核进入胚囊，一个精核和单倍体卵细胞受精，产生二倍体合子；另一精核与两个极核融合，产生一个三倍体核，这就叫做**双受精**（double fertilization）（图 2-9）。双受精仅在显花植物中存在。

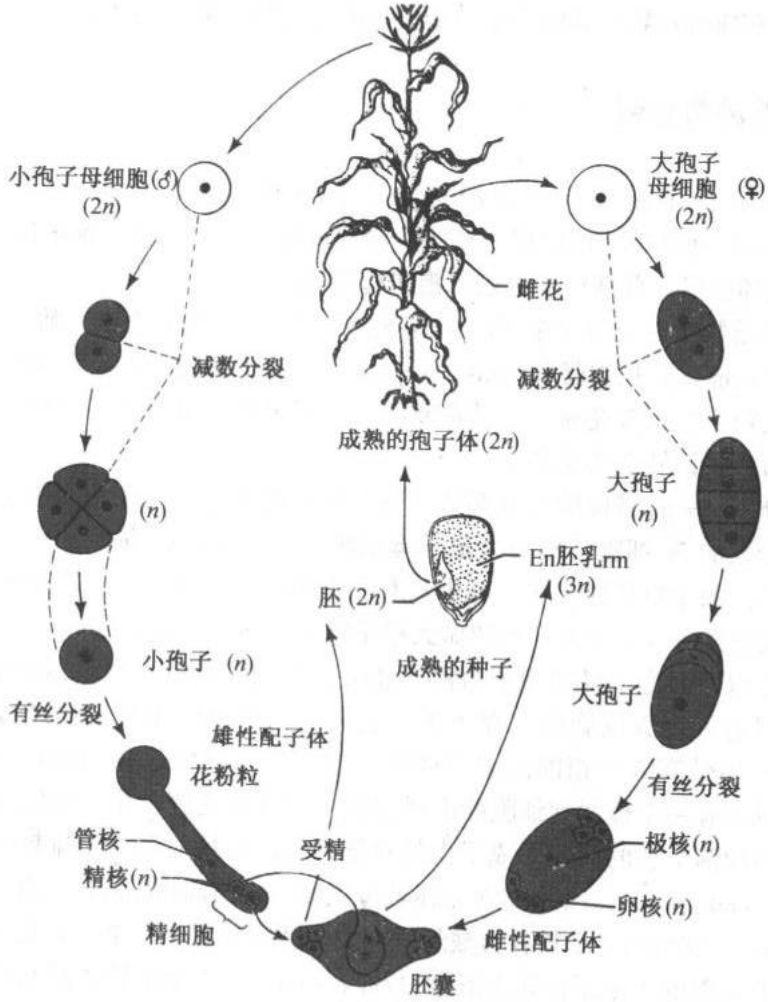


图 2－9　玉米的世代交替。(转引自 Griffiths A. J. F. *et al*. : An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3－40)

第三节　基因和染色体的关系

关于遗传的染色体学说，这个概念是把基因看作为染色体的一部分。这是由一位美国年轻的毕业生 Walter Sutton 和一位德国的生物学家 T. Boveri 提出的。1902 年孟德尔定律重新发现不久，引起人们的极大兴趣，他们两位独立地认识到豌豆产生配子时孟德尔因子的行为和减数分裂中的染色体行为有着精确的平行关系。在真核生物中基因是成对存在的，人们称之为等位基因，而染色体也是成对存在的，称为同源染色体；在形成配子时，等位基因相互分离，分别进入不同的配子中，一对同源染色体在第一次减数分裂时也互相分开，移向细胞的两极；不同的基因在形成配子时，自由组合地进入配子，非同源染色体也是如此（图 2－10）。这两位研究者通过分析得到了相同的结论。

遗传学的重要功绩就是解释了生物学中两大矛盾的生命现象“遗传”和“变异”，而有丝分裂和减数分裂提出了一个重要的线索：有丝分裂是一个维持遗传结构的过程，而减数分裂却是产生变异的过程，通过独立分配和交换而达到重组。

现代学习遗传学的学生们也许认为染色体学说并无惊人之处，但在 20 世纪初

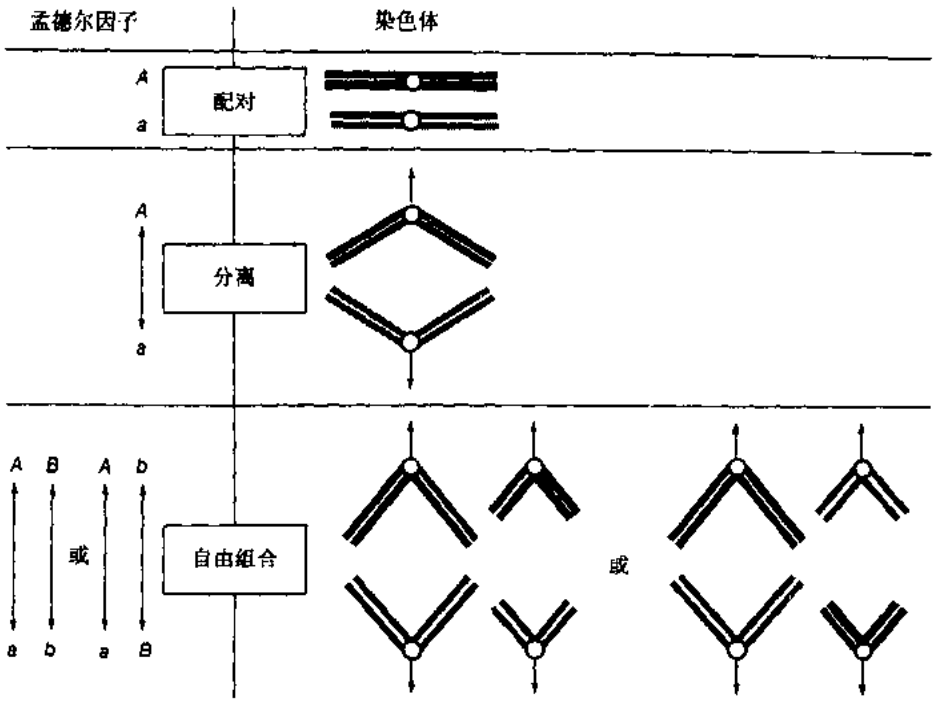


图 2-10 孟德尔基因和减数分裂中染色体的平行行为。深色表示一对同源染色体中的一个成员，浅色表示另一个成员。(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-6)

Sutton 和 Boveri 的假说却十分轰动，对此学说发表的第一个反应当然是寻找它的漏洞，几年之后终于引发了一场激烈的争论。对 Sutton-Boveri 学说提出的某些异议是有价值的，例如在当时，间期染色体是观察不到的。Boveri 对间期前后的染色体进行了细微的研究，试图回答一个问题，那就是通过间期后是否仍保持了完整结构。在有的生物中，有几对染色体看起来很相似，很难说它们的配对不是随机的。但在孟德尔定律中等位基因是有序的分离。然而在有的物种中染色体的大小和形态是不同的，它们可以用来证实在减数分裂中同源染色体的配对、联会和分离。

1917 年女细胞学家卡罗瑟斯（Carothers E）发现了一种染色体异常情况，在一种蚱蜢中，可用一种方法直接测出不同染色体配对时是否真的独立分配。通过研究她发现有一对染色体两个成员形态不同，此叫**异形配对**（heteromorphic pair），推测此对染色体是部分同源的。接着她又发现和异形配对无关的另一种染色体，仅单独的一条没有与其配对的染色体。她利用这些异常的染色体作为分离时染色体行为的细胞学标记。通过后期核的观察，她能对移向同一极的异形配对的不同染色体和没有“伴侣”的染色体进行计数（图 2-11）。她发现染色体分离有两种不同方式，但频率相同，虽然这些异源染色体的发现并不典型，但结果表明在配子中非同源染色体是自由组合的。

另一些研究者认为：由于所有的染色体都呈细线状，它们之间看不出有什么差别，几乎是各种染色体或多或少都是相同的材料构成的。这一相反的观点导致了一项突出的研究成果。1924 年布莱克斯利（Blakeslee, A.）对曼陀罗（*Datura stramonium*）的染色体开展研究，曼陀罗属茄科，有 12 对染色体，他发现有 12 个不同的品系，每个品系都有正常的 12 对染色体，但还多了一条额外的染色体，有的品系是第

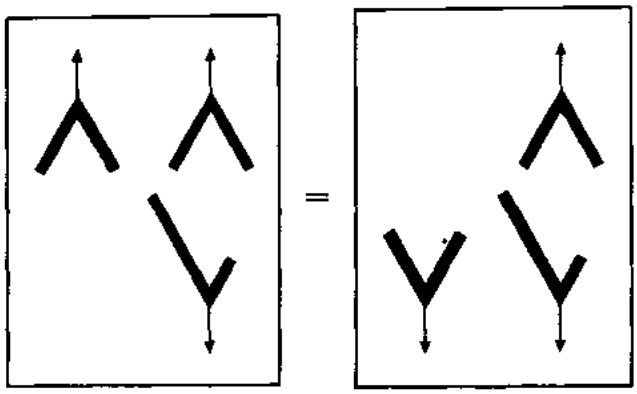


图 2－11　卡罗瑟斯发现异型染色体有两种分离的情况，但频率相同。
（转引自 Griffiths A. J. F. *et al.*：An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3－6）

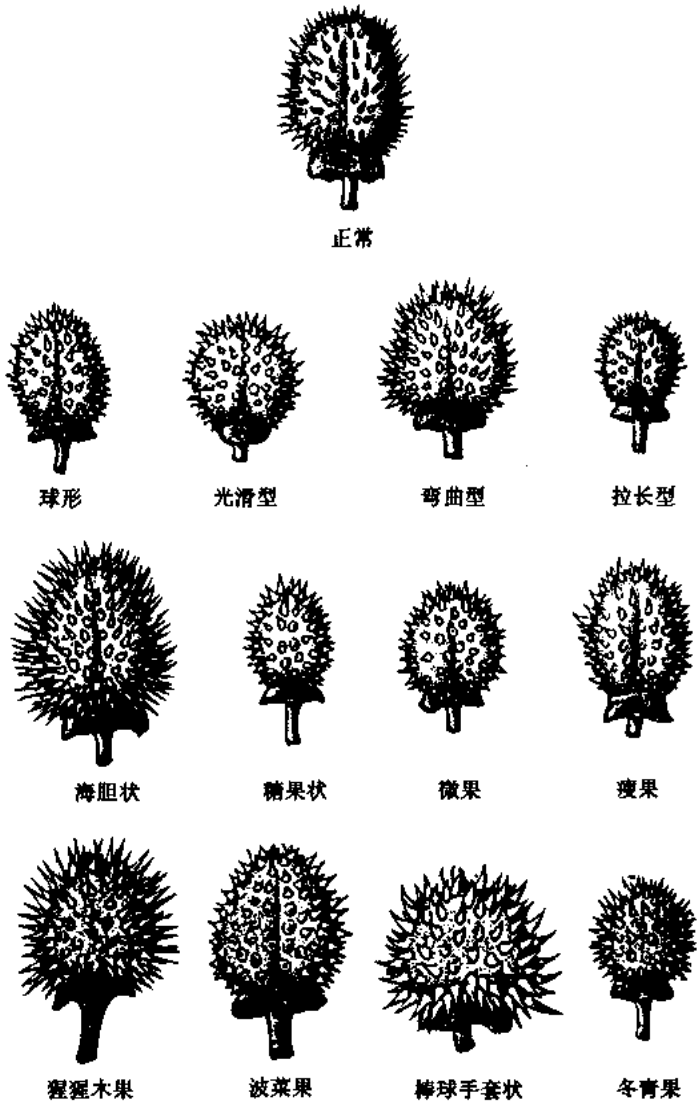


图 2－12　曼陀罗（*Datura stramonium*）有 12 个不同的品系，每个品系都有正常的 12 对染色体和一条不同的额外染色体，它们的表型也各不相同。（转引自 Griffiths A. J. F. *et al.*：An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3－8）

一对多了一条，有的品系是第二对多了一条，总之 12 个品系的染色体是不同的。布莱克斯利观察到 12 个品系表型不同（图 2－12），这个结果表明不同的额外染色体遗传学的效应是不同的。

这些结果都证实染色体的行为和基因的行为是完全平行的。使得遗传的染色体学说更具魅力，但还没有直接的证据表明基因位于染色体上，直到发现性连锁才提供了直接的证据。

第四节　性染色体的发现

遗传的染色体学说的证据来自于这样的实验，一些特殊基因的遗传行为和性染色体（sex chromosome）传递的关系。性染色体在高等真核生物的两种性别中是不同的。性染色体的发现为 Sutton-Boveri 的学说提供了一个实验证据。

在孟德尔定律第二次发现以前（1891 年）德国的细胞学家亨金（Henking, H）曾经用半翅目的昆虫蝽做实验，发现减数分裂中雄体细胞中含 11 对染色体和一条不配对的单条染色体，在第一次减数分裂时，它移向一极，亨金无以为名，就称其为“X”染色体。后来在其他物种的雄体中也发现了 X 染色体。

1900 年麦克朗（McClung, CE）等就发现了决定性别的染色体。他们采用的材料多为蚱蜢和其他直翅目昆虫。1902 年麦克朗发现了一种特殊的染色体，称为副染色体（accessory chromosome）。在受精时，它决定昆虫的性别。1906 年威尔逊（Wilson, EB）观察到另一种半翅目昆虫(Proteror)的雌体有 7 对染色体，而雄性只有 6 对，另外加一条不配对的染色体，威尔逊称其为 X 染色体，其实雌性是有一对性染色体，雄性为 XO 型。

在 1905 年斯蒂文斯（Stevens, N）发现拟步行虫属（*Tenebrio molitor*）中的一种甲虫雌雄个体的染色体数目是相同的，但在雄性中有一对是异源的，大小不同，其中有一条雌性中也有，但是是成对的；另一条雌性中怎么也找不到，斯蒂文斯就称之为 Y 染色体(图 2－13)。在黑腹果蝇中也发现了相同的情况，黑腹果蝇共有 4 对染色体，在雄性中有一对是异形的染色体(图 2－14)。在 1914 年塞勒(Seiler, J)证明了在雄蛾中染色体都是同形的，而在雌蛾中有一对异形染色体。他们根据异形染色体的存在和性别的相关性，发现了性染色体，现在已完全证实了他们的推论是完全正确的。

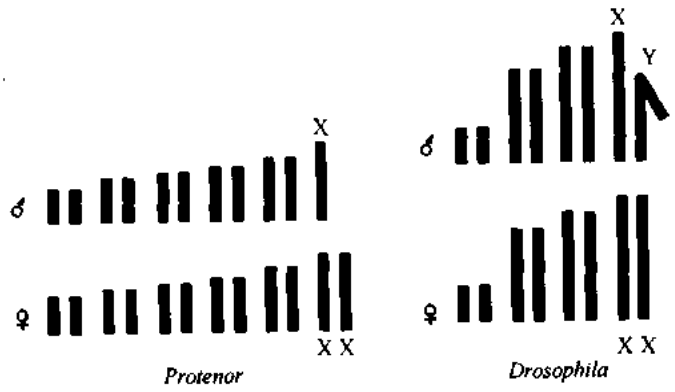


图 2－13　在两种昆虫中发现雌性和雄性的染色体有差异。

(转引自 Griffiths A. J. F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3－12)

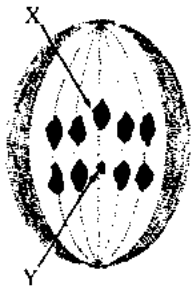


图 2－14　在步行虫减数分裂中异源配对的染色体（X 和 Y）相互分离。（转引自 Griffiths A.J.F. *et al.*：An Introduction to Genetic Analysis, 5th ed., 1993. Fig. 3－11）

严格地说异形染色体的存在仅是一条线索，而不是证据，不能因为存在异形染色体，就表明其为性染色体。一定要通过实验证明这条染色体上存在决定性别的主要基因，方能定论。

第五节　性连锁的发现

在 20 世纪初人们最初只发现互交的结果，产生相同的后代，并没有考虑到和性别之间有什么联系。第一个例外的情况是在 1906 年由唐卡斯特（Doncaster, L.）和雷纳（Raynor G H）发现的，他们研究鹅蛾（*Abraxas*）翅的颜色，用两种不同品系：一种是亮白色（dw），一种是暗白色（DW），正交表明亮白为隐性；反交结果表明翅的表型和性别有联系，雌性后代的表型和父亲相同，雄性后代的表型和母亲相同（图 2－15），如何解释这个结果呢？在解释前我们再考虑另一个例子。

♀亮白 × 暗白♂　　♀暗白 × 亮白♂

↓　　↓

暗白　　亮白(♀♀)　暗白(♂♂)

图 2－15　两种不同品系鹅蛾的正交和反交。

芦花♂× 非芦花♀　芦花♀× 非芦花♂

↓　　↓

芦花　　芦花♂♂　非芦花♀♀

图 2－16　鸡羽色的遗传。正交和后交结果不同。

贝特森研究了小鸡羽毛的遗传，一个品系是芦花的，另一品系是单一颜色（非芦花），不难看出，在正反交中和上例的结果一样。美国的科学家摩尔根（Morgan T H）根据自己的实验和敏锐的洞察力对以上的结果作出了令人信服的解释。他是位胚胎学家，最初（1903 年）摩尔根相信孟德尔的定律是正确的，不久由于他所进行的家鼠和野鼠的杂交实验结果显得毫无规律，使得他越来越怀疑孟德尔定律，但随着果蝇实验的成功，他不仅转变了自己的观点。而且逐步成为孟尔德学说最杰出的代表人物，1909 年他开始研究黑腹果蝇（*Drosophila melanogaster*）的遗传。果蝇易于培养，本身具有一些容易识别的性状，如眼色、体色、翅形、刚毛的形态等，生活周期短(图 2－17)，产生的后代多，这些优点使得果蝇作为一种遗传学研究材料要优于植物和高等动物。果蝇的应用是摩尔根成功的重要因素之一。

正常果蝇的眼色是红色的，摩尔根发现了一只白眼的雄果蝇，当他将这只雄果蝇和红眼雌果蝇杂交后所有 F_1 后代都是红眼，表明白色这个等位基因是隐性的。将 F_1 红眼果蝇互交，获得了 F_2 代，其红眼与白眼之比为 3∶1，但白眼都是雄的，红眼果蝇中雌雄之比为 2∶1，这是怎么一回事呢？

摩尔根为了收集更多的资料，又将雄性白眼与上述杂交产生 F_1 红眼雌性、红眼雄性、白眼雌性、白眼雄性四种后代。他又将第一组反交，结果性后代全为红眼，雄性后代全为白眼，这个结果和鸡羽毛及蛾翅的遗传非常相似，所不同的是在鸡和蛾子中，当雄性亲本带有隐性等位基因时，后代的性状和性别与亲代的情况正好相反，而在果蝇中，当雌性亲本带有隐性性状时，才出现这种情况。

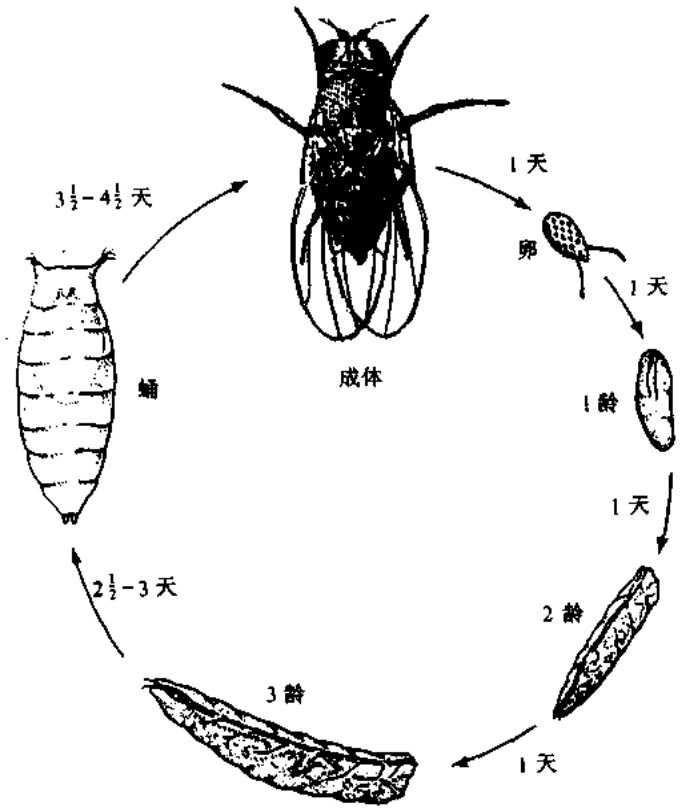


图 2-17 果蝇（*Drosophila melanogaster*）的生活周期。（转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-9）

摩尔根根据这些资料，首先假设 X 和 Y 染色体决定果蝇的性别，这样雌体在减数分裂产生的卵中都含有 X 染色体，雄性中虽然 X 和 Y 染色体异形，但也像同源染色体一样联会和分离，这样在雄性中可以产生两种类型的配子，分别含有 X 和 Y 染色体，而雌性只产生一种配子含 X 染色体，当含 X 染色体的精子和卵结合就产生 XX 型的合子，发育为雌性；当含 Y 染色体的精子和卵结合就产生 XY 型合子，发育成雄性。他的第二个假定是白眼和红眼是一定等位基因控制的性状，这对基因中的一个仅位于 X 染色体上，Y 染色体上是没有的，这样雌性有一对等位基因，而雄性只有一个。这样的大胆假设十分完美地解释了实验获得的各种结果。在白眼雄性和红眼雌性杂交中，F_1 代都是红眼，表明红眼等位基因是显性，我们可以用 W 来表示红眼，w 表示白眼，用 X^W 和 X^w 表示它们位于 X 染色体上，那么前面的杂交就可以

(a) 正交：

P 白眼♂ × 红眼♀

(X^wY) ↓ (X^+X^+)

F_1 红眼(X^+X^w, X^+Y)

↓ 互交

F_2 红眼♀ 红眼♂ 白眼♂

X^+X^+ X^+X^w X^+Y X^wY

2 : 1 : 1

(b) 反交：

P 白眼♀ × 红眼♂

(X^wX^w) ↓ (X^+Y)

红眼♀♀

X^+X^w

(c) 回交：

红眼♀ × 白眼♂

(X^+X^w) ↓ (X^wY)

红眼♀	白眼♀	红眼♂	白眼♂
(X^+X^w)	(X^wX^w)	(X^+Y)	(X^wY)
129	88	132	86

图 2-18 白眼和红眼果蝇的（a）正交、（b）反交和(c)回交。

用图 2 - 18 表示。

有些生物，如哺乳动物、果蝇等，在雌性中两条性染色体是**同配性别**（homogametic sex）；在雄性中两条性染色体是异形的，称为**异配性别**（heterogametic sex）。在另一类生物，如在鸟类和鳞翅目昆虫中，相反的雄性中有两条同形的性染色体，为同配性别；雌性为异配性别，用摩尔根的假设也可以得到同样完满的解释。

摩尔根的解释虽然发现了性连锁这样一个重要的规律，但仍是一种推理，还不能为 Sutton - Boveri 学说作为直接的依据。

第六节　染色体学说的直接证据

1916 年摩尔根的三大弟子之一的布里吉斯（Bridges，C）发现了 X 染色体的不分离现象，才找到了染色体学说的直接证据。布里吉斯将白眼（X^wX^w）雌蝇和红眼（X^wY）雄蝇杂交时大部分后代是红眼雌蝇、白眼雄蝇，这与摩尔根实验的结果完全相同，但有 1/ 2000 的后代出现了意外的情况：红眼不育的雄蝇和白眼可育的雌蝇。他称其为初级例外后代（primary exceptional progeny）。他又进一步把初级例外的雌蝇和正常红眼雄蝇进行杂交，结果约 4% 的后代是白眼雌蝇和可育的红眼雄蝇（图 2 - 19），他称其为次级例外，这是怎样发生的呢？

布里吉斯假设可能在雌蝇减数分裂时存在 X^w 染色体不分离的现象，同趋于一极，那么有一个子细胞中将有 2 条 X^w 染色体，而另一个子细胞中却没有 X 染色体；而雄蝇都能正常地减数分裂，这种异常的卵和正常的精子结合的结果有 4 种可能（图 2 - 20）$X^wX^wX^+$、单个 Y、单个 X^w♂ 和 X^wX^wY♀，前二者不能成活，后两者即是红眼雄蝇和白眼雌蝇，这就是次级例外。通过细胞学研究，果然红眼雄蝇只有一条 X 染色体，而白眼雌蝇有两条 X 染色体和另一条 Y 染色体，验证了布里吉斯的假设。

布里吉斯对唯一可育的初级例外（XXY）的减数分裂又提出了推理（图 2 - 21），分析有三种不同的分离类型，有 84% 是 X - X 染色体配对，然后相互分离，Y

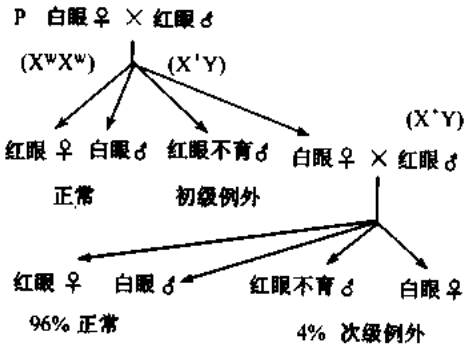
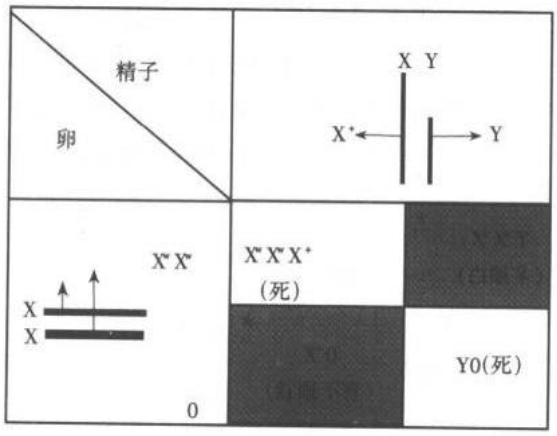


图 2 - 19　果蝇眼色遗传的初级例外和次级例外。（转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3 - 9）



★初级例外

图 2 - 20　由于 X 染色体不分离而产生了眼色遗传的初级例外。（转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3 - 17）

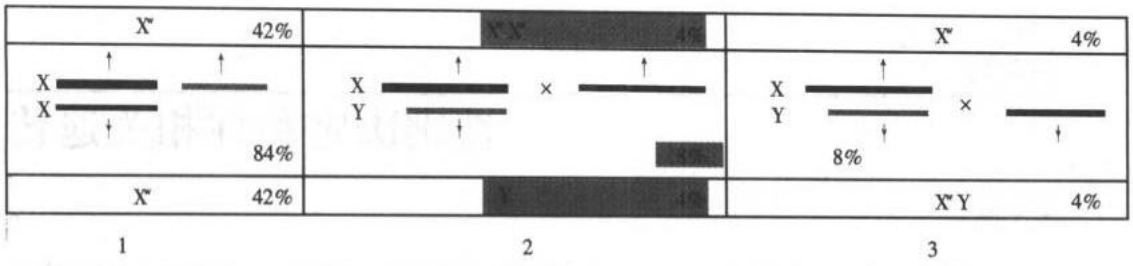


图 2-21 在 XXY 雄性果蝇中三种不同的分离方式。
(仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-18)

染色体随机地趋向一极；有 16% 的细胞是 X-Y 染色体配对，然后相互分离，游离的 X 染色体也随机地移向两极，因此有一半（8%）将形成 X Y 和 X 型的子细胞。

按这个假设初级例外形成的 6 种配子和正常的红眼雄蝇产生的两种精子杂交，将产生 92% 的正常后代和 4% 的次级例外(图 2-22)。这是预料的情况，实验的结果获得了从

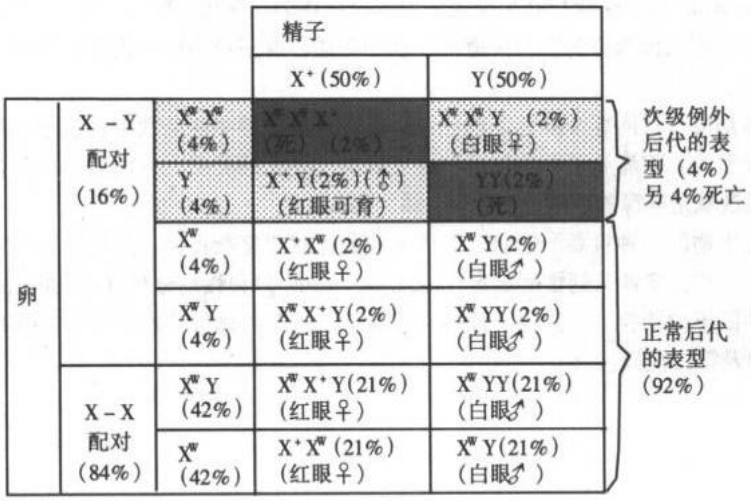


图 2-22 XXY 亲代产生特殊配子是形成次级例外的原因。(仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-19)

红眼雌蝇 27 679 只 白眼雄蝇 28 887 只 95.7%

白眼雌蝇 1 224 只 红眼雄蝇 1 246 只 4.3%

比例上看完全符合预料的结果，到此为止布里吉斯的模型比其他模型更具有说服力，他假设基因 *w* 或 *W* 在 X 染色体上，而且很好地解释了初级例外和次级例外，他的模型是受到了精确的检验的：

(1)初级后代的细胞学研究表明雌性为 XXY，雄性为 XO，证实了布里吉斯的推论；

(2)次级后代的细胞学研究表明雌性为 XXY，雄性为 XY，和推理相符；

(3)例外白眼雌蝇的红眼女儿一半为 XXY，一半为 XX，和镜检结果一致；

(4)例外白眼雌蝇的白眼儿子中也将产生例外的后代，这些白眼儿子都是 XYY，此也同样得到了证实。

布里吉斯的实验最终将 W/W^+ 基因定位在 X 染色体上，为遗传的染色体学说提供了有力而直接的证据，使遗传学向前迈出了重要的一步。

第三章

性别决定和性相关遗传

基本概念

初级性别决定是指性腺的形成。这可能受到特定基因的作用的结果。性别的分化是指在性腺所产生的性激素作用下，体细胞分化发育成内、外生殖器以及第二性征。

在单细胞真核生物中具有简单的性别决定系统，在此系统中由单对等位基因决定性别。在膜翅目昆虫中由一对杂合的染色体来决定雌性的发育。

不同的生物性别决定的机制不同，可分为四类，决定它们性别的主要因素分别是：①性染色体；②基因型；③性指数；④环境条件如温度。但归根结底都和基因的调节有关。

三种不同的性染色体系统分别是 XX－XO，XX－XY 和 ZZ－ZW。具有两个相同性染色体的性别称同配性别，带有不同性染色体的称异配性别。

大部分的植物都是两性的，少数植物具有 XX－XY 性染色体系统，称为雌雄异株植物。在女娄菜属（*Melandryum*）中性别是由 X 染色体和 Y 染色体的比来决定的。

和性别相关的遗传都称为性相关遗传。性相关遗传有三种不同的类型：伴性遗传、从性遗传和限性遗传。

性染色体是由一对同源常染色体进化而来。在有的物种中两条性染色体是同形的，有的物种性染色体是异形的。前者属于原始的类型，后者属于进化的类型。在两条异形的性染色体之间只有很短的一段区域是同源的，称为假常染色体同源区。

性别也是生物的一种重要的性状，但比一般的性状要复杂得多，从远古到现在性别一直是人们探索的热点，它主要涉及到**性别决定**（sex determination）和**性别分化**（sex differentiation）。前者指的是在各种因子的调控下，决定了性腺。后者指的是某些体细胞在性激素的作用下分化发育成内外生殖器以及第二性征。

第一节　生物的性别决定

一、性染色体

和其他性状一样决定性别的不外乎是基因和环境两大因素，在有的生物中基因的作用是占主导地位，几乎不受外环境的影响；而另一些生物看来环境因素似乎起着主导的作用。实际上环境仍然是通过基因在起作用。由于基因位于染色体上，因此在细胞学上所反映的常常是性别和染色体之间的密切相关。和性别决定相关的染色体被称之为**性染色体**。

在一般情况下性染色体的是由一对常染色体进化而来的，在进化时这一过程须要进行很多次。蛇就是一个很好的例子。人们研究了一些原始类型和高度进化的蛇的染色体。在原始类型的蛇中性染色体是相同的，称为同形（homomorphic）性染色体。这两条染色体的差异仅仅在于复制的时期不同（在 S 期的晚期），实际上是 W

染色体。在较进化的种中，其中一条 W 染色体变小了，形成**异形**（heterochromatic）性染色体，这就是 **Z 染色体**。**新莱昂（Lyon）假说**认为：性染色体是同形性染色体部分易位到同源染色体上所至，所以 Y 染色体小于 X 染色体，而 X 染色体往往有重复的基因（图 3－1）。低等脊椎动物的雄性和雌性体个体的性染色体在形态上彼此是不能区分的。在它们的性染色体的某些区段上，有许多等位基因的位点是两性所共有的，而在另一些区段上，两性所带的遗传信息却互不相同，雄性个体的性染色所携带的遗传信息主要与性别的决定有关。在鱼类中性染色体的分化就反映了这一进化趋势。

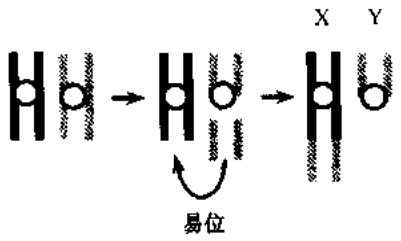


图 3－1 X,Y 染色体的形成。
（参考 Lyon M.F.：1974,Nature,250,651－653）

鱼类性染色体的分化可分为三种类型：①无性染色体的分化，性别由多个基因控制，影响性别分化的基因分散在多条染色体上，个体性别的发育决定于这些基因的平衡以及内外环境条件，每个基因的作用较小。因此，有些鱼类随着栖息条件和遗传型的改变，易出现性比的变化。如刨尾鱼（*Xiphophorus helleri*）及花鳉科的 *Limia vittata* 和 *L. caudofasicata* 就属于这种类型。②有性染色体的分化，但属同形。在性染色体上仅含有 1 个或少量专一性的性别决定相关基因，此是一类较完善的性别调节的类型。如花鳉科的某些胎生种类（*Poecilia reticulatus*, *P. variatus*）和鳉科的青鳉（*Oryzias latipes*），现借助于色素基因已成功地标记出性染色体。③有异形的性染色体。这是性染色体分化更为完善的一类。在一些极不相同的类群中如刺鱼和鲨鱼等也发现有异形性染色体，多数为雄性异配性别（XY 型），少数为雌异配性别（ZW 型）。Ojima 1983 年统计的数据表明在真骨鱼中发现有异形性染色体的不过 30 种。

二、性别决定

性别决定的本质是一些相关基因的表达和调控来决定的，这方面的内容将在发育（见第 24 章）中予以讨论，现在还是从染色体和环境的角度来加以阐明。

我们可以把性别决定初步分为以下三类。

1. 性染色体决定性别

在这类中我们又可以分成为 4 种不同的情况：

（1）性染色体的本身来决定性别

XY 型　在人类，哺乳动物，大部分的两栖类，爬行类，部分的鱼类和昆虫都属于 XY 型的。雄性为异配性别 XY，雌性为同配性别 XX。在这一类型中 Y 染色体起主导作用，不论 X 染色体有几条，只要存在 1 条 Y 染色体就发育为雄性。1990 年辛克莱尔（Sinclair,AH）等在前人工作的基础上发现在人和小鼠 Y 染色体的短臂上存在着**性别决定基因**，并在真兽亚纲动物中显示保守性。根据其在染色体上的位置，命名为 *SRY*（sex-determining region of Y chromosoine），近年来又克隆出一系列与性别分化有关的基因，但 *SRY/sry* 是其中起主导作用的基因，因此携带此基因的

Y 染色体成为决定雄性性别的标志。

在低等椎脊动物鸟类，两栖类，爬行类，鱼类以及某些植物中虽然也是由 Y 染色体来决定性别的，但机制尚不清楚。

ZW 型　　和 XY 型相反，以鸟类，鳞翅目昆虫及部分两栖爬行类等动物中是雌性异配性别（ZW），雄性同配性别（ZZ），但性别决定的机制不像 XY 型那样研究得比较清楚。按照一般的推测，W 染色体上可能也携带有和雌性发育有关的基因或带有抑制雄性发育的基因。

(2)性染色体的数目决定性别

在某些双翅目，直翅目和鳞翅目的昆虫中没有异形的性染色体，而是由性染色的数来决定性别。比如雌蝗虫：$2n = 24(XX)$，雄性为 $2n = 23(XO)$，比雌性少了一条染色体，称为 XO 型。植物中花椒也是属于 XO 型。在鳞翅目昆虫中也有雄性为 ZZ，雌性为 ZO 的类型，称为 ZO 型。

(3) 性指数决定性别

果蝇虽然也有 X 和 Y 染色体，但是其性别决定的机制和哺乳动物不同，不是取决于 Y 染色体是否存在，而是取决于**性指数**（sex index），即性染色体（X）和常染色体组数 A 的比。其分子机制已研究得比较清楚，我们在第二十四章中将予以讨论。线虫（*C. elegans*）也是由性指数来决定性别的。现在我们通过人类与果蝇性别决定的比较（表 3-1）就可以认识到果蝇性别决定的特点，其 Y 染色体并不决定雄性的发育，现已弄清果蝇的性别决定是由性染色体上一些基因的产物和常染色体上另一些基因产物的剂量比例来启动性别决定的开关基因，经过一系列的调控，来决定性别的分化（见第 24 章）。

表 3-1　在人类和果蝇中性染色体和性别的关系

	XY	XX	XXX	XXY	XO	XYY	X : 3A
性指数	1X : 2A = 0.5	2X : 2A = 1	3X : 2A = 1.5	2X : 2A = 1	1X : 2A = 0.5	1X : 2A = 0.5	1X : 3A = 0.33
人的性别	♂	♀	超雌	♂	♀	超雄	-
果蝇的性别	♂	♀	超雌(不能成活)	♀	♂	♂	超雄

(4)染色体组的倍性决定性别

蜜蜂(*Apis mellifera*)的性别决定十分特殊，是由染色体组的倍性决定的。蜂皇是可育的雌蜂，染色体为 2n = 32 条，经正常减数分裂产生的卵为单倍体 n = 16，卵和精子（n = 16）结合又形成 2n = 32 的合子，将发育成蜂皇和工蜂（2n = 32）。每一群蜜蜂中只有一个蜂皇，工蜂在遗传结构上和蜂皇并无差别，但由于工蜂所吃的蜂皇浆在质和量上都远比蜂皇差，所以发育成不可育的工蜂。蜂皇产生的部分单倍体的卵，少数未经过受精但也能发育，长成雄蜂（n = 16），雄蜂的减数分裂十分特殊，第一次减数分裂时出现**单极纺锤体**（monopolar spindle）所以染色体全部向一极移动，两个子细胞中一个正常，含 16 个二联体，而另一个是无核的细胞质芽体。正常的子细胞经第二次减数分裂产生两个单倍体(n = 16)的精细胞，发育成精子。

(5)取决于 X 染色体的是否杂合

小茧蜂（*Habrobracon*）的性别决定和蜜蜂又有不同之处，在自然状态下小茧蜂和蜜蜂相似，二倍体（2n = 20）为雌蜂，单倍体（n = 10）为雄蜂。在实验室中，人

们获得了二倍体（2n = 20）的雄蜂，其性别决定取决于性染色体 X 是纯合型还是杂合型。性染色体 X 有三种不同的类型：Xa、Xb、Xc。雌性的性染色体为一对杂合型的 X 染色体：XaXb 或 XaXc 或 XbXc；雄性为纯合型：XaXa、XbXb、XcXc。

(6)取决于 X 和 Y 的比

大部分植物是没有性染色体的，多为雌雄同株，有的为雌雄同花，少数植物为 XY 型，称为雌雄异株植物。女娄菜的性别是由 X 染色体和 Y 染色体的比来决定的。酸模（*Rumex acefosa*）的性别决定又属另一种类型，雌雄同株的植株为 18A + XX + YY，X∶Y = 1∶1；雌雄异株中雌株为 18A + XX + Y，X∶Y = 2∶1；雄性植株为 18A + X + YY，X∶Y = 1∶2。

但特殊的是，这两种交配型可以相互转换。具体机制尚不清楚。

2. 基因决定性别

(1)由复等位基因决定性别

植物中有异形性染色体的并不多，但有的植物却明显由基因控制的，如和葫芦科的一种喷瓜（*Ecballium elaterium*），是由复等位基因决定的(表 3-2)。

表 3-2 喷瓜的性别决定

基因和显隐性关系	决定性别	基因型
a^D	♂	a^Da^D、a^Da^+、a^Da^d
a^+	两性	a^+a^+、a^+a^d
a^d	♀	a^da^d

(2)由二对基因决定

玉米（*Zea mays*）一般都是雌雄同株的农作物。雌花的花序在叶腋呈穗状，由显性基因 Ba 控制的，其隐性等位基因为 *ba*(barren 的前两个字母)。雄花的花序在顶端称天花，由显性基因下 *Ts*(tassels)控制的，其基因型和表型性别的关系见表 3-3。

从表中可见 *Ba* 只控制叶腋是否长雌花序。而 *Ts* 基因则不同，在它为显性时顶端长出雄花序，在其为隐性时不论顶端长出雌花序。也就是说不论另一对基因的状态如何 *tsts* 存在时为雌性，*TsTs* 存在时有顶端开雄花。

表 3-3 玉米的性别决定

基因型	性别	表型
BaBa TsTs	♂♀	顶端长雄花序，叶腋长雌花序
Ba - tsts	♀	顶端和叶腋都长雌花序
baba Ts -	♂	顶端长雄花序，叶腋不长花序
baba tsts	♀	顶端长雌花序，叶腋不长花序

(3)环境决定性别

在表型性别（phenotynic sex)决定系统中，有时环境起主要的作用。典型的例子是海生蠕虫后螠（*Bonellia*），幼虫时没有性别的分化，自由独自在海底生活的幼虫发

育成雌性，长约 2 英寸，有一条很长的吻部，其他的幼虫附着在雌体中，寄居到雌体的子宫内，成为名副其实的寄生虫。这种螨虫的性别不是在受精时由遗传成分所决定的，而是由是否落在吻上而决定（图 3－2）。

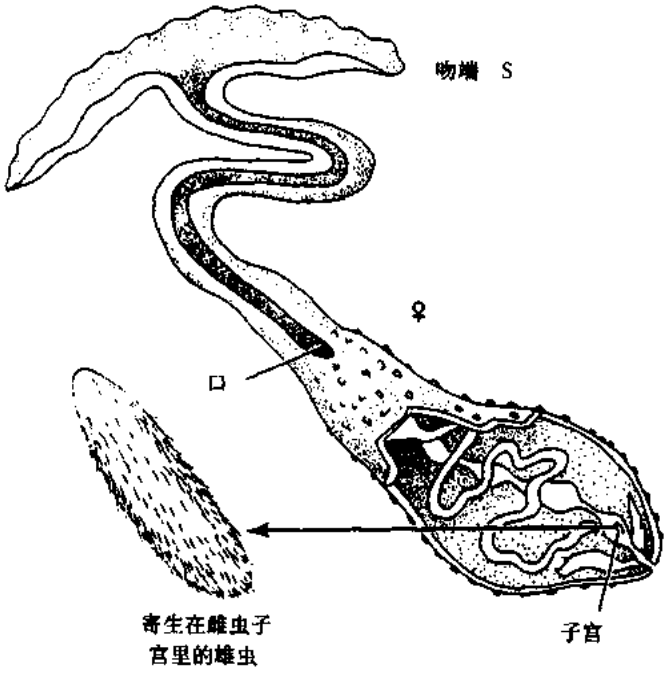


图 3－2　雌性和雄性的螨虫后螯（*Bonellia*）。
（引自 Russell, P. J.: Genetics, 3 rd ed., 1992, Fig. 3－18）

第二节　外界条件对性别发育的影响

一、外部的因素如温度、激素，感染等

两栖类中如蛙是有性染色体的，为 XY 型。蝌蚪在 20℃时 性别按其性染色体的构成正常分化。但在 30℃时，不论其基因型为 XY 还是 XX，都发育为雄蛙。在爬行类中也是如此。

如扬子鳄和密西西比鳄的卵在不同的温度下，发育为不同的性别，当在 30℃和 30℃以下，发育为雌体；当温度在 34℃或 34℃以上发育为雄体，乌龟的卵在 23℃～27℃的温度下发育为雄性，在 32℃～33℃时发育为雌性。所以他们卵产在不同高度的海岸线或河岸时由于温度不同而影响性别的分化，有可能是由于温度对两栖和爬行动物雌、雄激素的合成有直接的影响。

在葫芦科的植物里，丰富的氮肥，短的日照和夜晚的低温也有利于雌性的发育。

二、激素的影响

在哺乳动物中性别决定是指性腺的分化，性别的分化是指在性激素的作用下内

外性器官的分化和第二性征的发育，毫无疑问性激素的作用是很大的。

在双胎牛中如果是一雌一雄，由于雄性的睾丸先分化，睾丸产生的雄性激素通过绒毛膜血管，流向雌性胎牛，使雌性胎牛的内外生殖器表现为雄性，但无睾丸，呈间性。在人类中常见一男一女的异卵双生子中为什么没有这种现象呢？那是由于人类异卵双生是有两个胎盘将胎儿分隔开的，而牛的两个胎盘是共通的，由绒毛膜血管相连，所以可以相互影响。

三、性反转的现象

生物从一种性别转变成另一种性别的现象称为**性反转的现象**（sex reversal）。这种情况有的是正常的自然性反转，有的是人工引起的，或是环境条件改变所造成，或由于病理因素所引起。

酵母交配型的转变也属于性反转的现象，它是由基因控制的。在果蝇中存在一种**性别转换基因**(transformer)位于第 3 对染色体上。当 tra 雌果蝇的基因纯合时雌体发育成为不育的雄性。自然性反转在鱼类中有所发生，有一种海产鱼叫羊头原鲷，幼小时鱼体呈粉红色，为雄性，发育长大后变成黑色，性别转变为雌性。黄鳝（*Monopterus albus*），属合鳃目，在我国仅一属一种。黄鳝染色体 2n = 24，皆为端着丝点染色体，未发现异形性染色体的存在。黄鳝的体长和其发育阶段有着一定的相关性，在个体发育过程中产生自然性反转。2^+龄前皆为雌性，3^+ 龄转变为雌雄间体，卵巢逐渐退化，精巢逐渐形成。6^+ 龄全部反转为雄性。在欧洲鳗鲡（*Anguilla anguilla*），红鲈鱼（*Sacura margafacd*）及石斑鱼等的生活史中也有性反转的现象。

有的自然性反转和上叙相反，由雄性逐渐变为雌性。如一种海产的蠕虫（*Ophrgtrocha Puerilis*）幼虫都是雄的，随着发育长大雄性生殖器管退化，而雌性生殖器官发育，比较“老”的个体都是雌的。但用人工的方法如饥饿，或在水中加入钾盐，或将其身体后部剪去又可以促使雌体变成雄体。

牡鸡司晨的例子时有发生，有的母鸡停止产卵后竟然逐渐变成雄鸡，并能和雌鸡交配，但其基因型并未改变。

在人类中也存在着性反转现象，如 46XX 的男性和 46XY 的女性，这是由于 *SRY* 基因以及相关基因的异常所引起的（见第 25 章）。

性反转不仅是表型的改变而且还涉及到性腺的改变。在生物中虽然有的有性染色体的分化，看来每一个体其性别是可以向雌、雄两个方向分化的，也就是具备向两个方向发育的基因调节体系，外部条件或生理、病理条件的改变无非是影响了基因的调节而导致性反转。这个领域涉及到真核基因的调控，组织的凋亡，激素的分泌，组织的分化和发育，是十分复杂又非常有趣的。它已成为研究的热点。

第三节　性相关遗传

和性别有相关性的遗传都称为**性相关遗传**（sex-related inheritance），包括**伴性遗传，限性遗传**和**从性遗传**，三者的特点是不同的。

一、伴性遗传

伴性遗传（sex-linked inheritance）是指控制性状的基因在性染色体上，其遗传方式称伴性遗传，其遗传的特点由性染色体遗传规律所决定，基因的表达常不受性激素的影响。

前面介绍的果蝇红眼和白眼的遗传就是伴性遗传，概括起来有3个特点：①正反交的结果不同；②后代性状的分布和性别有关；③常呈一种**绞花式遗传**（crisscross heritance），即有害基因常常由母亲传给儿子，再由儿子传给外孙女。

1. 人类的伴性遗传

人类中一些遗传性状或遗传病的基因位于X染色体上，由于Y染色体上只有少数的基因，而一般没有X染色体上相应的等位基因，因此这些基因将随着X染色体来传递。在女性中由于有两条X染色体，所以即使其中一条X染色体带有致病的基因，由于是隐性的而得不到表达，我们称这种外在表型正常而带有有害基因的个体为**携带者**（carrier）。在男性中只有一条X染色体，隐性的有害基因没有对应的显性等位基因的存在，故称为半合子（hemizygote），这些隐性基因是可以表达的，正由于这个原因显示出绞花式遗传。男性的平均寿命低于女性可能与此有关。据调查**第一性比**，即男性胎儿和女性胎儿之比为120：100，可能由于含Y染色体的精子比含X染色体的精子更为活跃之故。**第二性比**是指男性婴儿与女性婴儿之比为（103～105）：100，仍然是男性多于女性，但已低于第一性比；可能是在胚胎发育过程中男婴的生命力比女婴弱；**第三性比**是指男青年与女青年之比为1：1，此表明人类的一夫一妻制的婚配制度正好是符合这一自然规律。也许是自然选择的结果。经统计85岁的老翁和老妪之比为62：100，相差悬殊，使人们不能不考虑到可能由于男性仅有一条X染色体的缘故。

（1）X－连锁隐性遗传

人类X－隐性遗传病已发现2000多种，最常见的是红绿色盲。较罕见是血友病，进行性肌营养不良，睾丸女性化和自毁容貌综合征等。

血友病（hemophilia）分为甲型和乙型两种，甲型是缺乏**第Ⅷ因子**，乙型是缺乏**第Ⅸ因子**，主要征状表现为凝血作用差，最著名的例子是英国维多利亚女王家族（图3－3）。

进行性肌营养不良（progressive muscular dystorphy，DMD）又叫做**假型肥大**（pseudomu-pertrophic）或杜兴氏症（Duchemme's muscular dystrophy）表现为肌肉无力，走路缓慢呈鸭步，易摔倒，上楼困难。仰卧起立时须先翻转为俯卧，再以双手支撑地面，然后再站起。常伴有腓肠肌、三角肌和冈下肌假性肥大。患儿一般在12岁左右发病，20岁左右死亡。这个基因DMD现已分离出，为根治这种遗传病带来希望。

图3－4是一个典型的X－隐性遗传家系图谱，从图中我们可以看出以下特点：①患者一般为男性；②有害基因由母亲传递；③呈绞花式遗传；④双亲正常，儿子可以是患者；⑤儿子若表型正常，后代都正常。

（2）X－连锁显性遗传（X-linked dominant inhertance）

由于控制性状的基因位于X染色体上，且为显性基因，女性有两条X染色体，

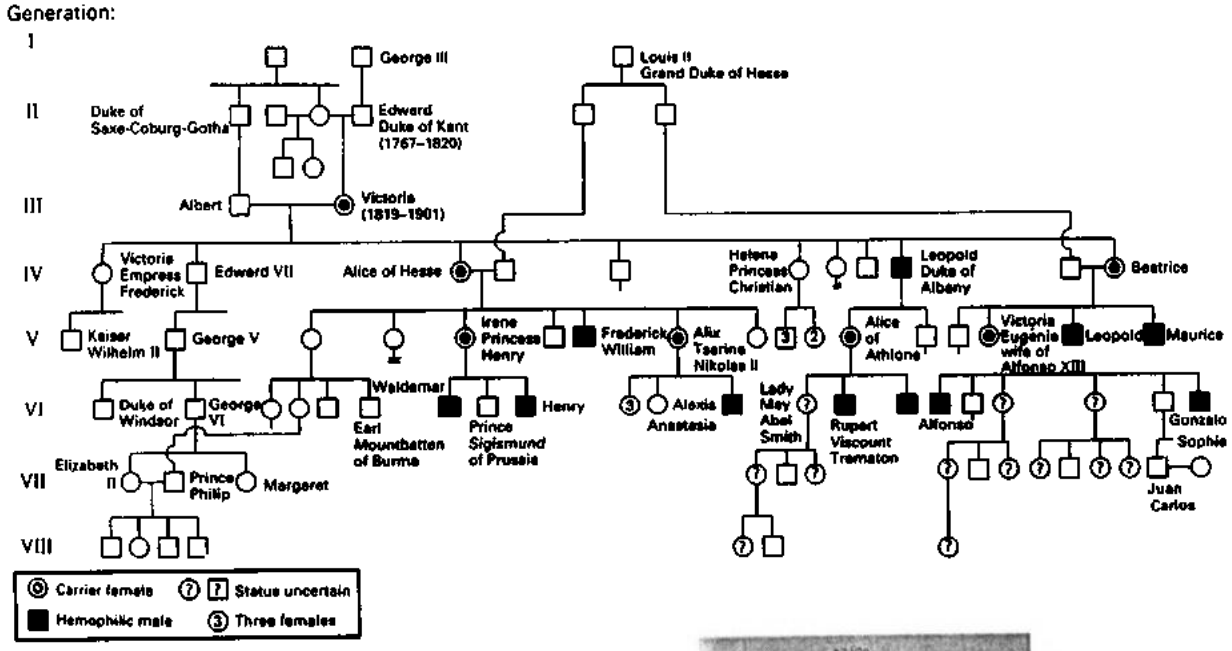


图 3-3 英国维多利亚女王(Queen Victoria)和其后裔的谱系表明血友病(属性连锁隐性遗传性状)的遗传规律。
(转引自 Russell, P.J.: Genetics 3rd ed., 1992, Fig. 3-19)

只要有其中任何一条 X 染色体有此基因都会出现相应的性状，而男性仅一条 X 染色体，带有此基因的概率仅是女性的一半。但女性患者多为杂合体，所以症状比男性为轻，如**抗维生素 D 佝偻病**(vitamin dresistant rickets)就是这类遗传病的一个例子。患儿由于肾小管不能对磷重新吸收，使血磷下降，尿里的磷增高，肠对磷、钙的吸收不良，因此形成佝偻病。

图 3-5 是个典型的抗维生态素 D 佝偻病的谱系。从图中可以看出：①患者女性多于男性；②每代都有患者；③男性患者的女儿都为患者；④女性患者的子女患病的机会为 1/2。

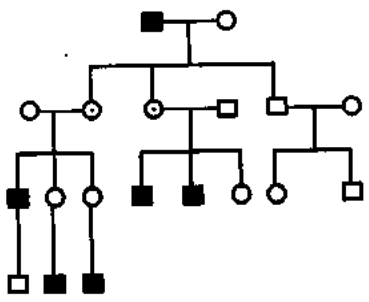


图 3-4 典型的 X-连锁隐性遗传家族图谱。

(3)Y-连锁遗传

Y 染色体上基因很少，而 X 上又没有相应的等位基因，因此无论是显性还是隐性都可以

C3-43
XJL

表达，直到现在人类中 Y－连锁惟一的例子就是**毛耳缘**（图 3－6），即成年男性外耳道中长出丛生硬毛，长约 2～3cm，伸于耳孔之外。在鱼类中 Y 连锁遗传的例子还是有的，如背鳍上的斑点就属于 Y 连锁遗传。

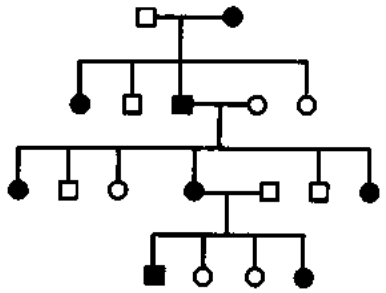


图 3－5　典型的 X－连锁显性遗传家族图谱。

2. 植物的伴性遗传

植物中有雌雄同花，雌雄异花，其中又有雌雄同株和雌雄异株之分，即使是雌雄异株的植物也并不都有异形性染色体，如杨柳等。少数植物有异形性染色体分化是少数植物，如女娄菜（*Melandrium album*）为雄性异配（表 3－4），草莓（*Fragaria*），金老梅（*Potentilla trutocosa*）为雌性异配。白剪秋萝（*Lychnis senno*）也是雄性异配性别的，叶片有

图 3－6　外耳道多毛的表型(a)(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3－31)和遗传谱系(b)。(转引自杜传书,刘祖洞主编:《医学遗传学》第二版,1991)

宽叶和窄叶两种，将宽叶雌株和窄叶雄株杂交，F_1 无论雌雄全部为宽叶，F_1 代互交后得以的 F_2 代雌性全部是宽叶，雄株半数为宽叶半数为窄叶，这个结果（图 3－7）和果蝇眼色的遗传也完全相同。不难推论，控制叶形的基因在 X 染色体上，而且 Y 染色体上没有相应的等位基因。

3. 鸟类的伴性遗传

鸟类伴性遗传最典型的例子就是鸡的芦花羽色的遗传，其规律和果蝇的眼色遗传也是相同的，不同的是鸟类为雌性异配，雄性同配而以。正交的情况如图 3－8 所示，反交不言而喻，和正交不同，具体如何，请读者自己推论一下，在此不再赘叙。

图 3－7　白剪秋萝的伴性遗传。

表 3-4 几种雄性异配性别的植物

物种	常染色体数	♀	♂
大麻(*Cannabis sativa*)	20	XX	XY
蔷草(*Humulus lupulus*)	20	XX	XY
酸模(*Rumex anglocarpus*)	14	XX	XY
女娄菜(*Melandrium album*)	22	XX	XY

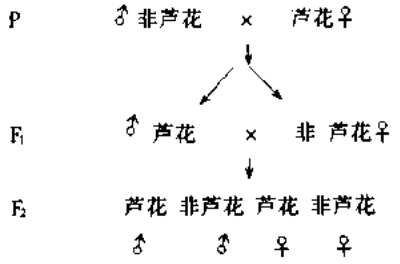


图 3-8 鸡羽色的伴性遗传

4. 不完全性连锁

在一对异形性染色体中，有部分区段是同源的，此段区域称为假常染色体同源区，减数分裂时可以联会。也有另一部分是不同源的区域(图 3-9)，在减数分裂时是不配对的区域。位于 X 染色体不同源区域上的基因和 Y 染色体不会产生交换，称为**X 连锁（X-linkage）基因**，同样位于 Y 染色体上的非同源区的基因就称为 **Y 连锁(Y-limkage)**。位于 X 染色体和 Y 染色体同源区域的基因称“**X 和 Y 连锁**”基因，由于这一区域有交换重组的可能存在，所以也称为**不完全性连锁基因**。如斯特恩（stern）发现果蝇的刚毛有一种突变型，刚毛变短，称截刚毛(bobbed)，它的遗传情况如图 3-10所示，其正反交的结果不是 F_1 不同，而是 F_2 不同，只能推论在 Y 染色体上有相应等位基因存在才可得到很好地解释。这也是惟一 X 和 Y 连锁的例子。

二、从性遗传

从性遗传是指基因在常染色体上，但由于受到性激素的作用，基因在不同性别中表达不同。如人类头发的早秃(Baldness)这一性状就属于**从性显性**(sex-influenced dominance)遗传。有的男子年纪很轻就开始秃顶，人们常误以为这是天资聪慧的标志，其实是一种遗传缺陷(图 3-11)。带有这个基因(*B*)的杂合体只在男性中出现早秃的表型，

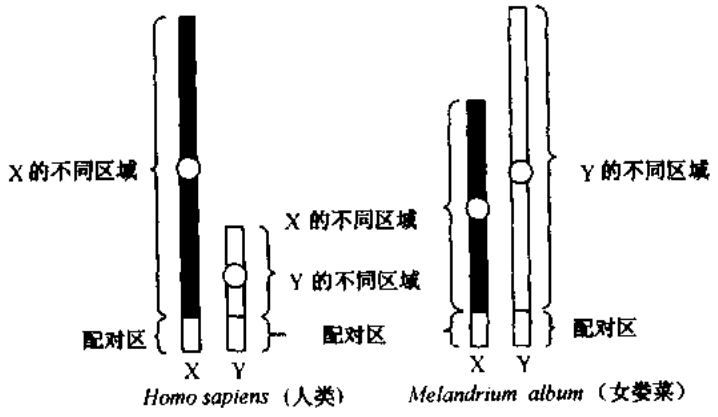


图 3-9 人类和女娄菜(*Melandrium album*)性染色体的配对区和不同源的部分。
(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3-21)

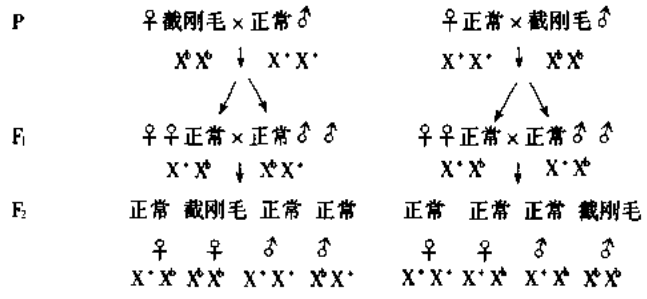


图 3－10　果蝇截刚毛的"X 和 Y 连锁"遗传方式。

在女性中表型正常，只有当 *B* 基因纯合时，女性才出现早秃，这个概率比带有单个 *B* 基因的概率小得很多，所以很少女性是早秃的。绵羊是否长角的情况和以上相似。鸡羽是一个很好的从性隐性遗传的例子，雄鸡的羽毛和雌鸡的不同，*h* 基因 控制雌鸡羽毛，只有当 *h* 基因纯合时在雄鸡中才能长出雄羽(表 3－5)。从图谱中(图 3－12)可以看出和 X－连锁不同，不是由 X 染色体传递的，而是由常染色体的传递，另外表现出一定的基因型有时在雌性中不能表达，而在雄性都得到表达，可能是雄性激素的作用结果。其本质是基因的表达受到内环境作用的结果。

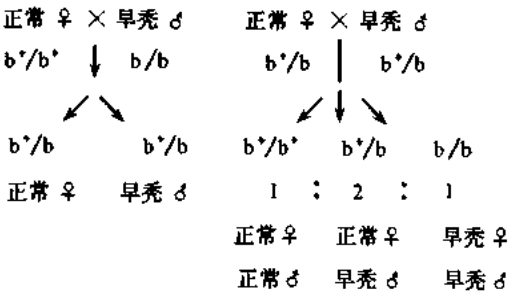


图 3－11　早秃的遗传方式。

三、限制遗传

限制遗传 (sex-limited inheritance)的基因可在常染色体或性染色体上，但仅在一种性别中表达，限性遗传的性状常和第二性征或性激素有关。

如毛耳的基因在 Y 染色体上，仅在男性中表达。睾丸女性化的基因位于 X 染色体上，但也只在男性中出现这种症状。子宫阴道积水也是一种遗传病，基因在常染色体上，但只在女性中出现。

表 3－5　从性遗传

类型	范例	基因型	表型	
			♀	♂
从性显性	人类早秃	b^+b^+	正常	正常
		b^+b	正常	早秃
		bb	早秃	早秃
从性显性	绵羊长角	h^+h^+	有角	有角
		h^+h	无角	有角
		hh	无角	无角
从性隐性	鸡羽形态	h^+h^+	雌羽	雌羽
		h^+h	雌羽	雌羽
		hh	雌羽	雄羽

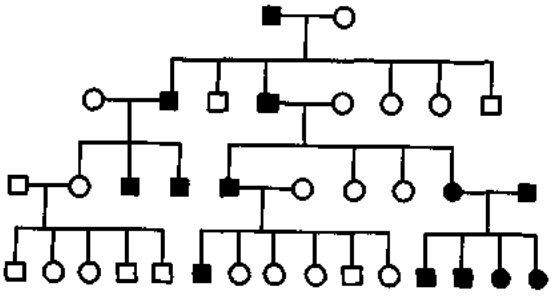


图 3－12　早秃的遗传谱系。

第四节 人类的性染色体异常

人类在减数分裂或在胚胎发育早期的有丝分裂中偶然会产生性染色体的不分离或丢失而造成了合子的性染色体异常，这些个体也将会出现性别分化的异常。

一、真两性畸形

真两性畸形不仅具有两性特征的内外生殖器，而且同时具有卵巢和睾丸两种不同的性腺，有 40%的患者一侧有卵巢，另一侧有睾丸；另 40%的患者一侧为卵巢或睾丸，另一侧为**卵睾**(ovotestis)；余下约 20% 的患者两侧均为卵睾。真两性畸形患者的副征性可为男性或女性，但其染色体却为嵌合型。主要分为两类：

(1) 46XX/47XXY；46XY/45XO；47XYY/45XO，这一类型是由于胚胎发育早期部分细胞有丝分裂过程中产生了**染色体不分离**的现象，(如 XYY/XO) 或**性染色体丢失**(如 46XXY/ 45XO) 所造成的，这些细胞最初来源一个受精卵，所以称为**同源嵌合体**(mosaic)。

(2) 46XY/46XX，这种类型十分罕见，可能是由双精受精的**嵌合体**(chimera)发育而成。人类的嵌合体因起源不同分为两类：**双胎嵌合体**(twin chimera)和**双精受精嵌合体**(dispermic chimera)。到 1975 年为止全世界发现上述两类嵌合体各 20 例。Dunsford (1953)首先发现双胎嵌合体，其特点是有**友伴双胎**存在，双胎之间由于交互移植，出现混合血液；双胎一方对另一方的移植的细胞缺乏凝集素，这是获得免疫耐受性的最好例子，这种嵌合体以自己的核型为主，他们具有正常的生育能力，我国在 1983 年也发现一例。Gartler(1962)发现另一类由双精受精引起的嵌合体。它是由两个精子和两卵核(可能为两个卵或一个双核卵，或一个卵，一个极体)受精后由于某种原因融合长成一个个体的结果，因而也称为**整体嵌合体** (whole body chimerism)，其特点是有混合型血液，体细胞培养物也显示镶嵌性。核型为 46XY/46XX，表现为真两性畸形，无生育力。

二、克兰费尔特综合征 (Klinefelter's syndrome)

又称为**"先天性睾丸发育不全"**或**"小睾症"**。核型为 47XXY，此是由于亲代减数分裂时，产生性染色体不分离所致，通过 X 染色体上 Xg 血型家系分析表现，60%的患者是由于母方，40%是由于父方染色体不分离所致。男性新生儿中的发病率为 1.3%，患者青春期后出现症状。身体瘦长，第二性征弱，常出现女性乳房，智力较差，睾丸小，精细管玻璃样变性，无精子发生，不育。

三、特纳氏综合征(Turner's syndrome)

又称为**"先天性性腺发育不全"**或**"先天性卵巢发育不全"**等。核型为 45XO，X 染色质阴性。1938 年由美国 H.Turner 首先报道。在女孩中的发病率为 0.2‰～0.4‰。患者表型为女性，第二性征发育不良，体矮(约 120～140cm)，常有蹼颈和肘外翻，卵巢缺如，原发性闭经，无生育能力。智力低下，常伴有先天性的心脏病。产生的原因也是由于

亲代配子形成过程中性染色体的不分离所致。约有 75%的丢失发生在父方。约有 10%的丢失了是在早期卵裂时。

Turner 综合征患者发病率远比 **Klinefelter 综合征**为低，这是由于约 98%的 45XO 胚胎在早期已流产。只有约 20%的个体发育异常较轻微才得以存活之故。

四、XXY 综合征

也称**超雄**，Sandberg(1961)首先报道，患者为男性，身材高大，少数有发育异常和智力较低，大部分核型为 47XYY，X 染色质阴性，有两个 Y 染色质，发病率占男性的 1/750 ~ 1/1500。1965 年 Jacob 等发现在暴力犯罪中 XYY 的比例较高（1：28），从而提出两个 Y 染色体的存在可能导致患者有侵犯行为。由于当时新闻界的渲染和扭曲，如把美国嗜杀成性的罪犯 Speek,R. 描述成超雄，更引起了混乱。然而，另一些研究者的报告指出 XYY 综合征患者的智力发育和社会行为完全正常，因此这种染色体异常对行为的影响尚无定论。

五、XXX 综合征

又称为**超雌**,Jacobs 等在 1959 年首先报道。发病率为 0.8‰,多数患者表型和生育能力都正常。只有少数患者继发闭经,智力稍低等症状,细胞学检查可发现患者有 2 个 Barr 小体。患者母亲的平均年龄增高,表明染色体不分离现象主要发生在母方。

第四章

孟德尔分析的扩展

基本概念

基因型要在一定的内外环境条件下才能表达为表型。环境对不同基因的表达的影响不同。

一个基因如果存在多种等位基因的形式，这种现象就称为复等位基因(multiple allelism)。任何一个二倍体个体只存在复等位基中的二个不同的等位基因。

在完全显性中，显性基因中纯合子和杂合子的表型相同。在不完显性中杂合子的表型是显性和隐性两种纯合子的中间状态。这是由于杂合子中的一个基因无功能，而另一个基因存在剂量效应所致。不完全显性中杂合体的表型是兼有显隐两种纯合子的表型。此是由于杂合子中一对等位基因都得到表达所致。

在很多情况下，由两对或多对基因通过对生化反应链不同步骤的控制，显示了非等位基因的互相作用。即一个基因的表型表达和另一对非等位基因有关，结果使孟德尔比发生变异。但对于二对基因来说，不论何种类型都是由 9∶3∶3∶1 的变化而来。

非等位基因之间的"显性"作用称为"上位"效应，"隐性"作用称为"下位"效应。某些基因的等位基因表达时可导致个体死亡，这就是致死基因。若杂合致死则为显性致死，若纯合致死则为隐性致死。致死基因可能还控制别的性状，其显、隐性作用和在致死中的显隐性作用有时并不一致。

一个生物的基因型决定其表型，但从基因型到表型要在一定的内外环境条件下通过发育才能实现。那么环境或多少对表型会有影响，同时表型是多种基因表型效应的共同作用的结果，基因间的相互作用又会对表型有什么影响呢？这一章将要讨论环境作用对表型效应的影响以及基因间关系对表型效应的影响。

第一节　环境的影响和基因表型效应

一、遗传和环境

环境条件对多基因遗传的影响是十分明显的，如人的身高、胖瘦、肤色、智商等性状是受多基因控制的，营养状况，生活环境，受教育的情况对这些性状都有直接的影响，即使是单基因遗传的质量性状也或多或少受到环境条件的影响。如半乳糖血症，是由于一个基因的突变使半乳糖苷酶缺失所致，但只要患儿不哺乳或不食用乳制品就不会患病。

水毛茛(*Ranunculus aguatilis*)的叶片形态在不同环境中是不同的。同一株植株，基因型完全一致，长在水下的叶片呈丝状，长在水面上的叶片呈掌状。

藏报春（*Primula sinensis*）有的品种在 20℃时花是红色的，而在 30℃时花为白色的。

喜马拉雅兔的毛色是受基因控制的在 25℃左右时在体温较低的部分，四肢和头部的尖端，尾巴和耳部的毛都是黑色的，其余部分全为白色。但在 30℃以上的环境里

长出的毛全为白色。如果把它躯干部除去一部分毛再放到 25℃以下的温度下，这部分新长出的毛是黑色的。当然生物有的遗传性状十分稳定，如人类的血型，也是受基因控制的，基因型一旦形成，表型就不会改变，不论生活在什么环境，营养状况如何，血型都不会改变。

二、反应规范

有一种类型的精神分裂症是由多基因控制的，具有相应遗传背景的人在平和的环境中并不一定会发病，但遇到某种精神刺激和心理压力常常会发病。因此这种类型的精神分裂症不仅和精神刺激有关，和遗传素质也有关。对于基因型正常的人来说常可以耐受一定程度的精神刺激，而基因型异常的人在相同的刺激阈值中就可导致精神分裂。图 4－1 可以表明基因型、环境和表型三者的关系。

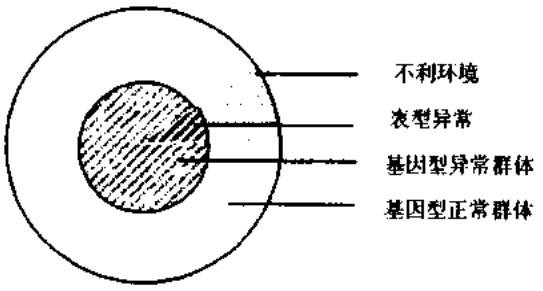


图 4－1　基因型、环境和表型三者间的关系。

1. 反应规范

我们怎样来衡量基因型，环境和表型三者之间的关系呢？对于一个特定的基因型而言，我们可以制一个表格来表示在不同的环境中基因通过发育将产生什么样的表型。对于特定的基因型而言，这样一套环境－表型的相关性就称为这种基因型的**反应规范**（*norm of reaction*）。也可以说是某一基因型不同环境中所显示出的表型变化范围。例如果蝇在不同的温度下复眼发育大小也不同。在这一物种中我们就可以用各种不同复眼数基因型来测果蝇的反应规范。如有三个基因型：野生型，中棒眼和小棒眼（图 4－2a），在不同的温度下发育，然后来计算单眼的数目，绘成曲线（如图 4－2），曲线就显示出三种反应规范。野生型果蝇在 15℃～30℃环境中，复眼是从 1000 逐步减少到 700。中棒眼随着温度的升高，复眼反而增加，小棒眼的在同样的环境下，复眼减少，虽然减少的数目比野生型小，但幅变减少得比野生型大。

2. 表现度与外显率

表现度（*expressivity*）是指在一定的环境中，某一突变个体基因型表达的变异程度。例如多指是由显性基因控制的，带有一个有害基因的人都会出现多指，但多出的这一手指有的很长，有的很短，甚至有的仅有一个小小突起，表明都有一定的表型效应，但变异程度不同。如克汀病患者，同样患病，但症状有轻有重，重者生活无法自理，轻者可以从事简单劳动，可以做 10 以内的加减法。这就是表现度的不同。

外显率（penetrance）指在特定环境中某一基因型（常指杂合子）显示出相应表型的频率（以百分比表示）。也就是说同样的基因型在一定的环境中有的个体表达了，而有

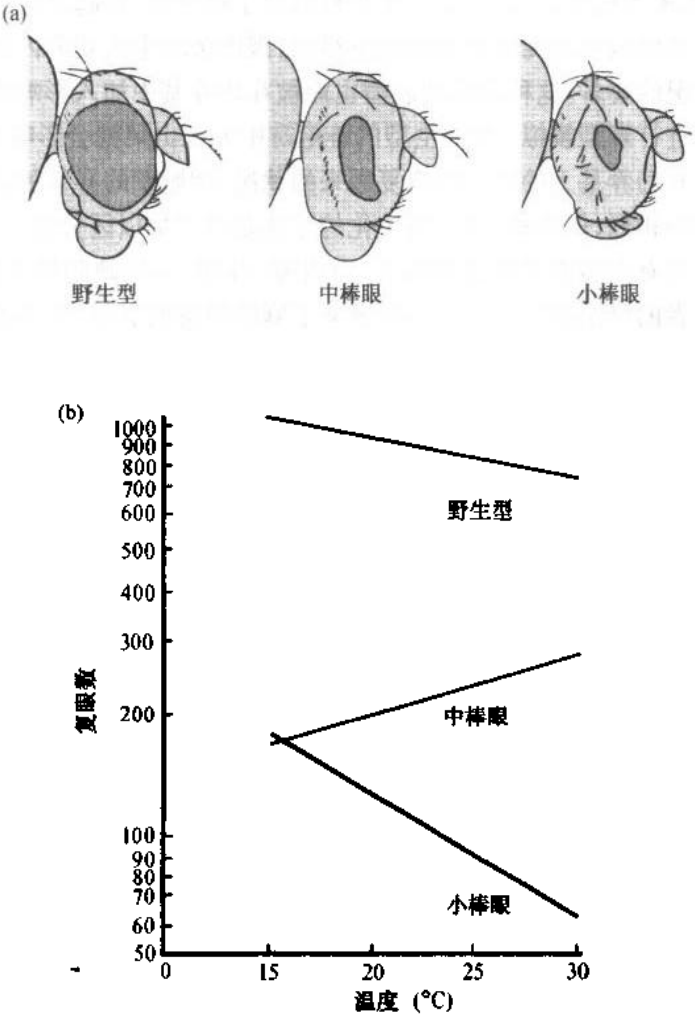


图 4-2 果蝇中控制复眼数的三种不同基因型(a),对不同温度的反应规范(b)。
(转引自 Griffiths A. J. F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 1-13)

的个体未得到表达,表型正常,如具某显性基因 A 的杂合体(Aa)有 100 人,其中仅有 40 人出现相应的表型,则外显率为 40%,如 100 人都出现相应的性状,则外显率 = 1,称**完全外显**,若外显率 < 1 称**不完全外显**。

3. 拟表型或表型模拟

环境的改变也能影响一个或多个基因的表达,最敏感的时期是在发育早期,在这个时期较小的变化可以导致以后较大的变化。**表型模拟**的定义是:用特殊环境条件下所作的一个非遗传的表型的修饰,使其和一个突变型的表型相仿。产生表型模拟的药剂称**表型模拟剂**(phenocopying agent)。表型模拟的例子很多。这种表型模拟的异常个体在遗传结构上并没有任何改变。

表型模拟存在各种生物中,从简单到复杂的生物都可找到一些例子。如红色链孢霉(*Neurospora crassa*)正常的生长表型是有很多分枝。在网状团块中的菌丝很快地延伸、长成桔黄色,与灰霉或绿酶相似,有时长在面包上,有些突变的链孢霉分枝很多,而

长得很慢，称为**菌落样**生长习性。由于生长慢形成了接种点，看起来像是固体培养基上的一颗纽扣。其细胞壁结构缺陷使细胞壁很弱，因此菌丝中胞质的内部压力使得产生分枝比正常情况的要多。这种菌落的表型也能被在培养基中加入影响细胞壁正常发育的化学试剂而进行**表型模拟**。在野生型的链孢霉中加入山梨糖会影响菌落生长。若将此菌落放在正常培养基上的话它将恢复原来的状况。细胞壁的异常的表型模拟与菌落的基因突变表型相似，但本质不同，用生化的方法是可以加以区分的。

在人类中也有很多表型模拟的例子，白内障、耳聋、心脏缺陷等有的是遗传的，也有的是由于患者的母亲在怀孕后 2 周内感染了麻疹病毒而引起。人类的其他一些性状也有表型模拟，如**短肢畸形**(phcomelia)，这种畸形有的是受显性等位基因控制的，这个基因抑制了四肢长骨的发育。在 1959～1961 年之间，发现孕妇在妊娠 35～50 天若服用镇静剂**反应停**(thalidomide)也会使孩子产生同样的表型，现在这种药物早已被淘汰了。

最近发现很多得复杂有机体在一个发育的敏感时期受到表型模拟剂的影响会产生表型模拟。很多不同的药物都起到这种作用。

4. 双生子研究

我们怎样来区别一个性状是由基因控制的，还是由环境控制的呢？一种办法就是研究在不同环境下基因型相同的个体，如我们可以用同一基因型的实验动物。典型的就是将一定品系的小鼠或大鼠，通过近亲杂交产生的后代用作实验材料。它们大部分基因是纯合的。这个动物群体一般说来都是同型的。可用于一些未知药物的研究，如新的药物，食品添加剂，食品色素，防腐剂及染料等。通过动物实验来评估对人类是否有潜在的危险。

虽然我们不可能制造人类近亲婚配品系用于实验，但我们可研究双胞胎。双生子有**异卵双生**(fraternal twins)和**同卵双生子**(identical twins)。异卵双生是两个卵同时受精的结果，他们的基因型和同胞的兄弟姐妹一样，但他们的胚胎发育环境以及以后的生长环境是相同的。同卵双生子是单卵受精，第一次卵裂时分成两个细胞，然后各自发育成个体。他们在遗传结构上是完全相同的，但他们后天生长的环境可以不同。我们只要加以比较分析就能分辨某一性状是由环境决定的还是由基因决定的。如很多孩子都患过麻疹，无论是同卵双生子，还是异卵双生子同时患病率都是高的，这表明患麻疹在很大的程度是由环境决定的。有的性状在两种不同的双生子中同一性是有明显差异的，如脉搏的频率，血型及精神分裂的发病率在同卵双生子中是高度的一致性。资料(表 4－1)表明，肺结核在不同人之间的共发病率。肺结核是由结核杆菌所引起，但在群体中有人患病，有人不患病，说明不同的人易感性不同，这种易感性是否由基因控制呢，从资料中明显得出同卵双生子的同时发病率为 87.3%，而异卵双生子的同时发病率为 25.6%，差异是明显的；而异卵双生子和同胞之间的差别并不明显，说明肺结核的易感性和遗传有关。

表 4－1　不同情况下肺结核的共发病率

关系	同时患病率%
同卵双生子	87.3
异卵双生子	25.6
同胞	25.5
半同胞	16.9
亲子	11.9
夫妻	7.1
不相关的人	1.4

5. 性状的多基因决定

我们说质量性状是由单个基因控制、并不意味着产生某一性状仅由一对基因决定，其实任何一个性状都是经过一系列生化连锁反应的结果，在反应的过程中可能存在多个步骤，而每一步也都由相应的酶进行催化，这一系列的酶同样是由一系列基因控制合成的，因此每一个性状都涉及到多个基因。将突变型和野生型相比较并不是一系列基因同时都发生了突变，而仅仅是其中一个基因发生突变，使整个生化反应不能完成，从而不能表现出来正常的性状。我们指单个基因控制也就是指仅一个基因发生突变也会改变表型的性状。

比如哺乳动物皮毛的颜色就是一个好的例子。小鼠的毛色至少由 5 对基因来决定，即 *A*, *B*, *C*, *D* 和 *S*。野生型等位基因 *A* 控制的表型称野鼠色(agouti)，也就是在黑毛的前端产生一段黄色，从整个动物看起来像野兔子那种深棕色。其等位基因是不能产生黄斑，毛呈纯黑色，即 *aa* 鼠为黑色。*B* 基因是使色素元形成黑色素，其等位基因 *b* 是不能形成黑色素，毛呈纯黄色。*C* 基因是控制色素(color)的合成，而其等位基因 *c* 则色素不能形成，毛色为白色，各种白化的动物都是这个基因突变所致。有些动物 *C* 基因的表达同时受到环境条件湿度的影响，即在 30℃ 以上这个基因不表达，即使是野生型，也不能产生色素而呈白色，而在 25℃ 下才能表达，所以喜玛拉雅白化兔和泰国猫的尾巴是黑色，其他部位是白色。*D* 基因是起淡化（dilutes）颜色作用的基因，野生型 *D* 不起淡化作用，突变 *d* 可将颜色变淡，杂合型 *Dd* 起到中等淡化的作用。*P* 基因是控制动物皮毛上有无花斑，*P*: 控制无斑，*p*: 有斑；人们常见的花斑马、狗、猫、鼠等都是由这个基因控制的。那么一只纯黑色小鼠它的基因型如果用 5 个基因来表达的话应当是 *aaBBCCDDPP*。在 *a*、*C*、*D*、*P* 相同的情况下，*B* 为黑色，*bb* 为棕色，我们就说是 *B* 基因控制黑色和棕色，其实纯黑色是五对基因共同作用的结果。

6. 基因的多效应

一个基因的产物无论是酶还是其他的蛋白，它的生化功能或生理功能不可能是惟一的或独立的，常会影响到其他生化反应及生理功能，因此一个基因的突变引起的表型变化也是多方面的，这种现象我们称之为**多效现象**(pleiotropism)，多效现象的例子也是很多的，如人类的**先天性成骨不全**（osteogenesis inperfecta），这是种常染色体的显性遗传病，患者骨骼发育不良，骨质疏松，一生中反复发生多次骨折；大多数患者巩膜呈淡蓝色；患者成年后还常常出现耳聋。这样，一个座位上的基因发生改变，产生 3 种不相关的表现改变。人类的**指甲－髌骨综合征**(nail-petalla syndrome)也出现多效现象，此病属常染色体显性遗传。患者指甲异常。第一、二指指甲萎缩，角化不全，纵裂或缺如。髌骨缺如或发育不良，桡骨小头，肱骨髁不对称发育，脱位，髌骨畸形等。少数病人还有眼和精神方面的异常。

第二节 显隐性关系的扩展

孟德尔在植物杂交实验中所观察的 7 对性状都属于完全的显性和隐性关系。但并

不是所有情况都是如此。有时会遇到一些例子在显隐性关系上出现各种变异。

一、不完全显性

不完全显性（incomplete dominance）又叫做**半显性**（semidominance），其特点是杂合子表现为双亲的中间性状。如紫茉莉（*Mirabilis jalapa*），红花品系和白花品系杂交，F_1 代既不是红花，也不是白花，而是粉红色花；F_1 互交产生的 F_2 代有三种表型，红花，粉红花和白花，其比例为 1∶2∶1。金鱼草的花色也是如此。安大路西亚（西班牙南部一个区域）鸡的羽毛、家蚕的体色，马的皮毛，金鱼身体的透明度等（图 4－3）都属此类不完全显性。为什么会产生不完全显性的现象呢？以紫茉莉为例，两个正常 *R* 基因产生的酶的剂量才能产生足够的红色素。当基因发生无效突变时，便失去功能，不能催化红色素的产生、故 *rr* 为白色。只有一个正常的 *R* 基因，其产生的酶就只能产生部分红色素，所以 *RR* 为红色，*Rr* 为粉红色，*rr* 为白色（图 4－4）。马皮毛的颜色是因为 *D* 是淡化基因，马的棕色由 *bb* 决定 *DD* 不起淡化作用，*dd* 起很强的催化作用，使皮毛呈白色，*Dd* 只有一个 *d* 起淡化作用，使马呈淡棕色。同样具有剂量效应（图 4－3d）。

我国学者陈桢曾系统地研究了金鱼的起源和遗传。发现普通金鱼（*TT*）能合成酪氨酸氧化酶，使酪氨酸在细胞里合成各种色素，呈现出绚丽的色彩。有一种突变型（*tt*）是酪氨酸氧化酶缺陷型，不能合成色素颗粒，所以身体透明，从外面可以看到金鱼的内脏。普通金鱼和身体透明的这两种金鱼的 F_1 代是一种半透明鱼，也就是说单个“*T*”基因合成的色素量尚不能完全改变透明状态，所以杂合体呈半透明状态（图 4－3e）。

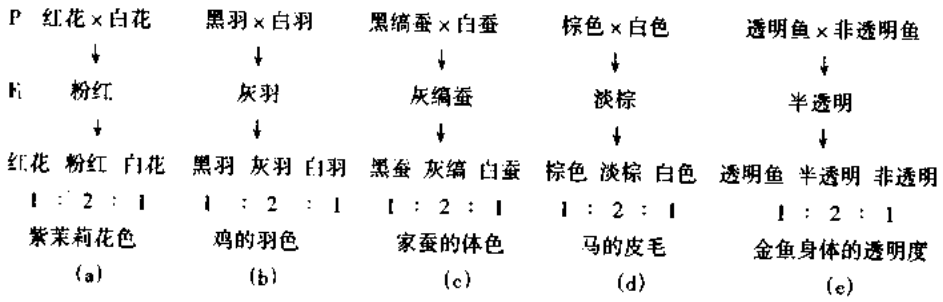


图 4－3　不完全显性的遗传方式。

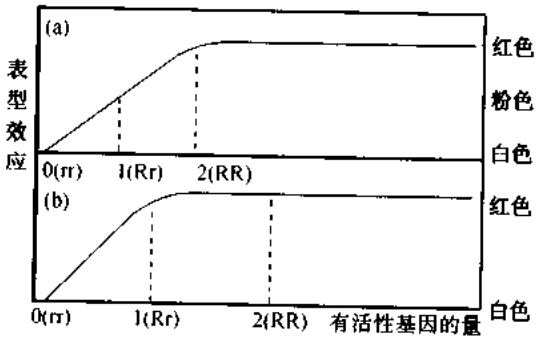


图 4－4　紫茉莉（*Mirabilis jalapa*）花色不完全显性的机制（a）和完全显性（b）花色的遗传之间的差别。（转引自 G.D.Elsetl: Genetics, P18）

二、共显性

杂合子的一对等位基因各自都具有自己的表型效应，称为**共显性**(codominance)。MN **血型**是最好的例子。在人类的 M－N 血型系统中有三种血型，M，N 和 MN 型，他们是由基因型 L^ML^M，L^NL^N 和 L^ML^N 决定的。M 型个体的红细胞膜上有 M 抗原，N 型个体的红细胞膜上有 N 抗源，而 MN 型个体的红细胞膜上既具有 M 抗原又有 N 抗原，也就是两种基因在同种组织中都得到了表达。这是由一对等位基因中的一个发生了异效突变，它会产生不同的表型效应，当这一对等位基因杂合时，两种表型（M 抗原和 N 抗原）同时共存。

三、嵌镶显性

我国遗传学家谈家桢先生早年系统地研究了鞘翅瓢虫（*Harmonia axyridis*）的遗传。发现了**嵌镶显性**(mosaic dominance)。一种瓢虫鞘翅的底色为黄色，前缘呈黑色，称为**黑缘型**（$S^{Av}S^{Av}$），另一种情况相反，鞘翅的后缘为黑色，称为**均色型**（S^ES^E），两者杂交产下的 F_1 代同时具有双亲的特点，即鞘翅前后缘都呈黑色，这就是嵌镶显性。它和共显性是不同的，共显性是在同一组织同一空间表现了双亲各自的特点，而嵌镶显性是在不同的部位分别表现了双亲的表型嵌镶在一起(图 4－5)。

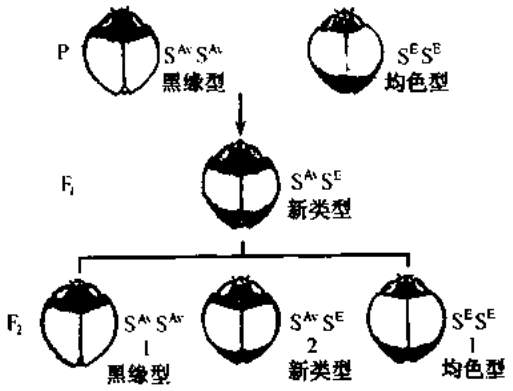


图 4－5　瓢虫鞘翅色斑的遗传。黑缘型与均色型杂交，子一代显示嵌镶遗传现象

四、显隐性的相对性

1. 标准不同显隐性不同

我们在讨论显隐性关系时，总要以某种性状为标准来分析，同一对等位基因若以不同的标准来讨论，那么显隐的关系可能是不同的，如人类的**镰刀型贫血**（sickle cell amemia）是由于珠蛋白 β 链上的第 6 个疏水性的氨基酸取代负电性的亲水性的谷氨酸所引起，使镰刀型血红蛋白 HbS 在脱氧状态下比正常血红蛋白 HbA 的溶解度低 5 倍，于是在氧张力低的毛细血管区，HbS 溶解度大大降低而形成管状结构凝胶化，导致红细胞成镰刀状，称为镰变。从下表 4－2 中可以看出标准不同显隐性也不同。

表 4－2　镰刀形贫血显隐性的相对性

基因型	临床表现	镰形细胞	镰形细胞数	含HbS量	电泳
$Hb^A\ Hb^A$	正常	无	无	0	HbA一条带
$Hb^A\ Hb^S$	正常	有	1/3	20%～40%	HbA、HbS两条带
$Hb^S\ Hb^S$	贫血	有	绝大多数	90%	HbS一条带
遗传规律	隐性	显	不完全显性	不完全显性	共显性

2. 不同环境条件对显隐性的影响

(1)外部环境

光照和温度对生化反应是有重要的影响，这样也必然会影响到表型效应，以致改变显隐性关系。如玄参科的金鱼草有红色花和淡黄色花两种不同的品系，若果将这两种不同品系进行杂交，所产生的 F_1 代在不同的条件下，表型不同(如图 4－6)。

红色花 × 淡黄色

↓

光充足低温：　红色花

光不足温暖：　淡黄色

光充足温暖：　粉红色

图 4－6　金鱼草有红色花和淡黄色花两种不同的品系
杂交后，所产生的 F_1 代在不同的条件下，表型不同。

在光充足低温时 F_1 为红色，那么红色为显性；当光不足温暖条件下，F_1 为淡黄色，那么红色为隐性；当光充足温暖时，F_1 为粉红色，呈不完全显性。可见外部的环境条件会影响到显隐性的关系。

(2)内部环境

前面介绍的早秃等也是属于内环境影响显隐性的例子。*Bb* 杂合的情况在男性中受到雄性激素的作用，单个的秃发基因可以表达，产生早秃性状，成为显性基因，而在女性中，由于没有雄性激素，所以不能表达，一定在有害基因纯合的情况才能表达，从而表现为隐性性状。

第三节　复等位基因

在遗传学发展的早期就已经清楚，一个基因可能有两种以上的形式，然而一个二倍体生物在一个基因座位上只可能具有其中一对等位基因，而在一种生物群体中，某一个基因座位上的不同等位基因的总数常是相当大的。这种情况称**复等位基因**(multiple allelism)，一套等位基因的本身就称为**等位系列**(allelic series)。**等位性**是遗传学中重要概念之一。所以我们要了解几个例子。

1. 人类的血型遗传

人类的血型系统共发现 24 种，其中最常用的是 ABO 系统。伯恩斯坦(Bernstein)对 ABO 血型的遗传提出了复等位基因的假设，此系统共由 3 个复等位基因 I^A，I^B 和 i 控

表 4-3 人类 ABO 系统血型表

血型表现	基因型	抗原(在细胞膜上)	抗体
A	I^AI^A, I^Ai	A	β(抗B)
B	I^BI^B, I^Bi	B	α(抗A)
AB	I^AI^B	AB	-
O	ii	-	αβ(抗A抗B)

制，I^A 和 I^B 互为共显性，但对 i 为显性。各种血型的基因型如表 4-3 所示。I^A 基因已定位在 9q34，控制合成 N-乙酰半乳糖转移酶，此酶能把 α-N-乙酰半乳糖分子接到蛋白上，产生 A 抗原决定簇；I^B 基因控制合成半乳糖转移酶，能把 α-N-半乳糖接到糖蛋白上，产生 B 抗原决定簇，i 基因为缺陷型，不能合成 A 抗原和 B 抗原。血型在法律上常用于亲子鉴定，排查嫌疑犯的依据，血型的遗传是十分稳定的，但也有一些特殊的情况，在分析时必须注意。

(1)顺式 AB(cis AB)型

有位 AB 型的妇女和 O 型的男子结婚，生育了 O 型的子女(如图 4-7)表面上看起来似乎是不符合血型遗传的规律，经仔细的分析发现一般正常情况下 I^A 和 I^B 是在一对同源染色体上，称为**反式 AB**(*trans* AB)，但也有极少数的人，由于交换重组 I^A 和 I^B 同位于一条染色体上，另一条同源染色体上没有任何等位基因。这种情况称为 *cis* AB。发生率为 0.18‰。当带有 *cis* AB 染色体的配子和 O 型人的配子结合，此配子的染色体带有三基因，故合子发育成 AB 血型，而 *cis* AB 的同源染色体进入配子，这个配子和正常 O 型血型的配子结合，合子仅能发育成 O 型(图 4-8)。

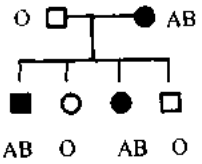


图 4-7 顺式 AB 的一个谱系。

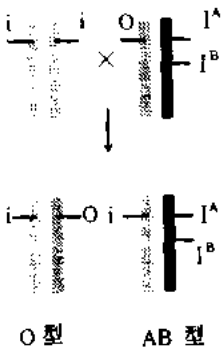


图 4-8 顺式 AB 产生的机制。

(2)孟买型

一个 O 型的男子和 B 型的女子结婚，生育了一个表型为 O 型的女孩，O 型女孩长大后和一个 A 型血男子结婚，生育了两个女儿，一个为 O 型，一个为 AB 型(图 4-9)。一个 O 型的和 A 型人结合，怎么会生育出 AB 型的子女呢？似乎又是违反了血型遗传的规律，实际上这是由于抗原的形成由多个基因控制的结果(如图 4-10)。前体在 H 基因控制下产生 H 抗原，即岩藻糖基转移酶，作为 A、B 抗原的前体。明白了这道理，就不难解释上述的现象。那就是Ⅱ-2 女子的"O"型是一个假象，她的"H"座位基因型可能为 hh，即不能产生 H 抗

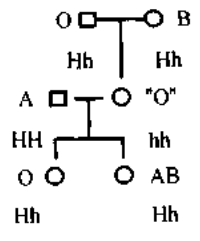


图 4-9 孟买型血型的谱系。

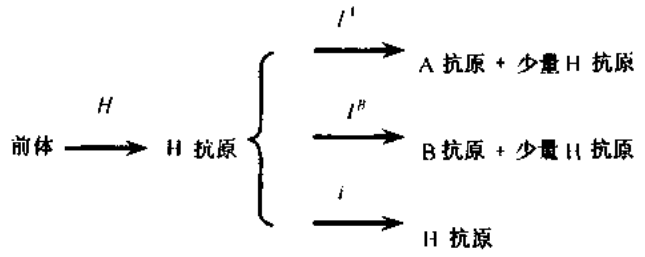


图 4 - 10　ABO 血型中 A 抗原和 B 抗原的合成。

原，所以即使其 ABO 系统为 I^B i 型，因没有前体物也不能形成 B 抗原，故定为"O"型。当和一个基因型为 *HH*/I^Ai 正常男性结婚，他们的孩子 *H* 座位的基因型为 *Hh*，能产生 H 抗原，ABO 座位的基因型可能是 I^Ai，I^Bi，I^AI^B，*ii*，当为 I^AI^B 和 *ii* 时为 AB 型和 O 型。这一情况是在印度孟买发现的，故称**孟买型**。

(3)HAB 的分泌型和非分泌型

HAB 不仅存在于红细胞的表面，还可以通过分泌系统转运到血清，唾液，胃液，精液经以及体液中，因此法医除了收集血样还可以通过唾液，精液等来确定血型，这种能分泌的称为**分泌型**(secretor)，也有的人不能分泌，称**非分泌型**(nonsecretor)。分泌型为显性，非分泌型为隐性。

(4)Rh 血型与新生儿溶血症

Rh 血型也是常用的一种血型系统，表型只有两种 Rh^+ 和 Rh^-，Rh 有 Rh^+抗原，其化学本质为粘多糖，基因型为 *RR* 和 *Rr*；Rh^-无 Rh 抗原，基因型为 *rr*。*R* 对 *r* 为显性，在欧洲和北美 Rh^-在群体中的比例为 15%，在中国人中仅有 1.5%。Rh^-的比例较高会给新生儿带来一种潜在的危险。当 Rh^-的女子和 Rh^+男子结婚，怀孕时胎儿为 Rh^+，胎儿产生的 Rh 抗原少量经胎盘进入母体。使母亲产生 Rh 抗体。由于母亲本身并无抗原，所以抗体的存在对其毫无影响。但抗体也同样可以通过胎盘进入胎儿体内，和胎儿的 Rh 抗原结合，引起溶血。特别是第二胎，母亲受到 2 个胎儿 Rh 抗原的刺激，相应产生 Rh 抗体的量就比较多。第二胎产生的溶血现象也就较为严重。假如母亲为 Rh^+，胎儿为 Rh^-时，则不会产生溶血，因为母亲体内产生的抗原进入胎儿体内，胎儿此时尚无产生抗体的能力，所以不会产生溶血反应(图 4 - 11)。

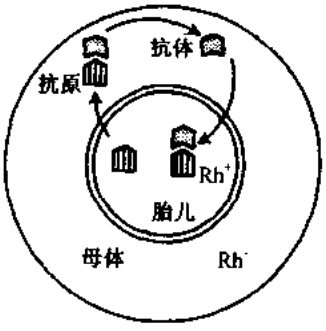


图 4 - 11　母亲为 Rh^-、胎儿为 Rh^+时产生新生儿溶血的机制。

2. 组织相容性抗原

人类在进行器官移植时会产生**异体排斥现象**，这是由于异体之间**组织相容性抗原**(majorhistocompatibility，MHC)或称人类白细胞抗原(humanleucocyteantigens，HLA)的差异所致。人类的 *MHC* 基因在第 6 号染色体上有 9 个座位(A、B、C、D、*DR*、*DQ*、*DP*、*DZ*、DC)，每个座位又由多个复等位基因构成。小鼠的 *MHC* 基因在第 17 号染色体上，有 6

个座位，*K*、*D*、*L*、*R* 基因编码移植抗原；*I* 编码 B 细胞和 T 细胞的表面蛋白；*S* 编码补体。就以人类的 4 个位点为例(表 4－4)基因型组合有 4 358 651 661 种，表型组合也有四亿多(434 194 112)如果以 9 个位点计算，数字更为庞大，所以除同卵双生子外，世界上很难找到 MHC 完全相同的人。但器官移植并不要求供体和受体的 MHC 完全一致，而要求供体的基因受体都具有(表 4－5)。当然这也是相当困难的。

表 4－4　人类 MHC 四个位点的基因数和表型数

	A	B	C	DR
复等位基因数	21	38	9	13
表型数	211	704	37	79

表 4－5　两个位点供受体基因的排斥反应

移植组	受体基因型	供体基因型	反应
1	$A_1A_2B_5B_8$	$A_3A_4B_5B_7$	排斥
2	$A_2A_3B_7B_{12}$	$A_1A_2B_7B_8$	排斥
3	$A_1A_2B_7B_8$	$A_1A_2B_7B_8$	接受
4	$A_2A_3B_7B_8$	$A_2A_3B_8B_8$	接受

3. 兔子毛色

兔子的毛色也是由一组复等位基因构成的(表 4－6,7)，C^+ 控制野鼠色，对其他 3个复等位基因呈显性。C^{ch} 控制灰色，对 C^h 和 C^a 呈显性；C^h 控制喜玛拉雅白化，对 C^a 为显性；C^a 是一种突变型，不能形成色素，纯合时兔子皮毛为白色。

表 4－6　兔子毛色的复等位基因

等位基因	基因型	表型
C^+	C^+C^+、C^+C^{ch}、C^+C^h、C^+C^a	野鼠色
C^{ch}	$C^{ch}C^{ch}$、$C^{ch}C^h$、$C^{ch}C^a$	灰色
C^h	C^hC^h、C^hC^a	喜马拉雅白化
C^a	C^aC^a	白化

复等位基因所组合成的基因型数可由通项公式来计算(表 4－7)。

表 4－7　一定数量等位基因构成基因型的数目

等位基因	基因型种类	纯合子种类	杂合子种类
1	1	1	0
2	3	2	1
3	6	3	3
4	10	4	6
5	15	5	10
n	n(n+1)/2	n	n(n－1)/2

4. 致死基因

正常的小鼠的皮毛是棕灰色的，1905 年法国学者居埃诺(Lucien Cuenot)研究小鼠时发现了一只黄色小鼠，他将这只黄色小鼠和纯系正常小鼠杂交产生的后代一半是黄鼠一半为正常的小鼠。再将黄鼠和黄鼠杂交，后代中 2/3 是黄鼠，1/3 是正常鼠。

为什么会出现这样的比例呢？从第一组杂交(1)看来黄鼠不是纯合的，否则后代中不会出现二种表型；(2)黄色对野生型为显性。第二组杂交的分离比不符合杂合子互交后的孟德尔分离比 1∶2∶1，而是 1∶2 。显然可能有一种显性基因纯合类型未能成活(图 4-12)。

这个假设和第一组杂交的结果也是吻合的。这种导致生物死亡的等位基因就称为**致死等位基因**(lethal allele)。一些基因当它们突变时会引起致死表型，那么这些基因就称为**必要基因**(essential gene)。小鼠的致死基因发现后，人们不断地在人类及动物中发现很多类似的例子。致死基因可以分**显性致死**(dominant lethal)和**隐性致死**(recessive lethal)为两类。像黄鼠这个例子，A^Y 基因在控制毛皮的颜色上是显性的，但在"致死"这个表型效应上必须是纯合状态下才能实现，也就是说是属于隐性的，因此我们可以称为隐性致死。人们发现在灰色皮毛的银狐中有一种变种，在灰色背景上出现白色的斑点，十分漂亮，称白斑银狐。用这种皮毛制成大衣可谓是"千金裘"。于是人们希望培育

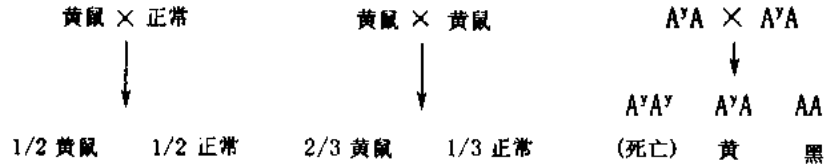


图 4-12　致死基因。

出白斑银狐的纯系，但将它们互交后育下的后代有两种表型：白斑银狐和银狐，比例为 2∶1，结果和黄鼠的情况如出一辙，说明控制白斑的基因纯合致死。曼岛猫(Manx cat)(M^tM) 是没有尾巴的，和上面的情况相似，没有纯系 (M^tM)，此是由于当基因型为 M^tM^t 时在胚期致死。在人类中很多隐性遗传病都会致死，如镰刀形贫血，半乳糖血症等，也属于隐性致死。另一种类型是显性致死，如人类**视网膜母细胞瘤**(retinolastoma)，为显性遗传。此是由于抗癌基因 *Rb* 发生突变或缺失而引起，患者多为 1～5 岁儿童，即使摘除眼球，成年时还会转发骨癌或肺癌而致死。

第四节　基因间的相互作用与修饰

任何一个表型都是由多对基因控制的，现在我们来讨论两对非等位基因怎样来控制同一性状？他们相互间的关系是怎样的。根据孟德尔定律两对基因杂交其 F_2 代的表型之比是 9∶3∶3∶1。前面讨论的是两个相互独立的性状，现在讨论的是同一个性状，显然情况不同，但万变不离其宗，仍然符合孟德尔的法则，产生一些新的表型只不过是一种表型的综合而以。

在等位基因中有显性和隐性之分，在两对非等位基因控制同一性状时也有"显"、"隐"之别，为了和等位基因的"显性"，"隐性"相区别，人们就把两对非等位基因中起"显性"作用的称**上位基因**(epistatic genes)，将起"隐性"作用的一对基因称为**下位基因**(hupostatic genes)或者说上位(epistasis)是指基因相互作用中一对基因的表型表达取决

于另一对非等位基因的基因型。也就是异位显性。

一、共上位

观赏鸟类中有一种叫虎皮鹦鹉的鸟，它的毛色有绿、蓝、黄、白四种，都是稳定遗传的，若将纯种的绿色和白色鸟杂交或将蓝色或黄色的鸟杂交，F_1 代是绿的。F_1 代进行互交，F_2 代羽色产生四种表型：绿、蓝、黄、白，比例为 9∶3∶3∶1(图 4-13)。假设蓝色素是 *B* 基因控制的，黄色负素是由 *Y* 基因控制的，*B* 和 *Y* 都是完全显性的话，将黄鸟和蓝鸟杂交得到的 BbYy 是绿色鸟，也就是黄色和蓝色的色素同时得到表达，混合在一起，我们看到的当然是绿色(图 4-14)，如果 *B* 基因和 *Y* 基因都发生突变而不能形成色素，那么 *bbyy* 应当是白色，这样的假设正好符合虎皮鹦鹉的杂交。等位基因中有"共显性"，那么在非等位基因中相似的现象我们就称为共上位(coepistatic)。其机理可能是如图 4-14 所示。

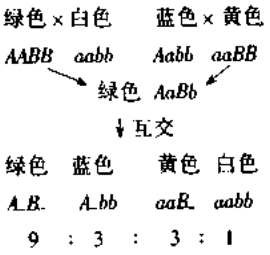


图 4-13 共上位效应。

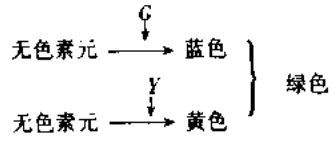


图 4-14 共上位效应的生化机制。

二、嵌镶上位 G 基因

前面已介绍了鼠皮毛的多基因控制，野鼠色的基因型为 *AABB*，*B* 基因使得黑色的毛的前端有一条黄色斑带，整个看起来是野鼠色，也就是深棕色。纯棕色鼠为 *aabb*，因无 *A* 基因，所以没有带(图 4-15)。黑鼠色小鼠和棕色鼠杂交或者黄鼠和黑色鼠杂交，F_1 代为野鼠色，F_1 互交后 F_2 代也产生四种表型(图 4-16)，野鼠色，黄色，黑色，棕色，它们的比例为 9∶3∶3∶1。此和共上位不同，当 *A* 和 *B* 都是显性时，虽同时表达，但空间部位不同，和嵌镶显性一样，我们称之为**嵌镶上位**(mosaic epistatic)。

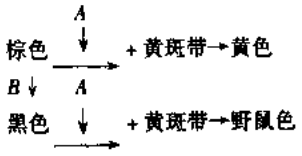


图 4-15 嵌镶上位产生的机制

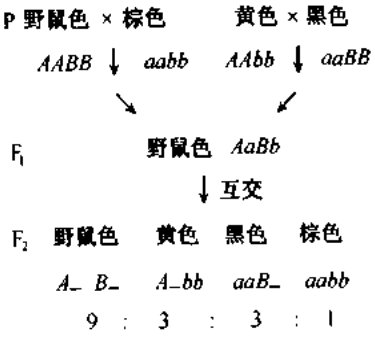


图 4-16 嵌镶上位效应。

三、半上位

如猪的毛色，有一种棕红色，基因型为 *AABB*；另一种为白色，基因型为 *aabb*，他们的 F_1 代仍是棕红色，F_2 代有三种表型，棕色，淡棕色和白色。比例为 9∶6∶1(图 4-17,18)，这个结果明显是 9∶3∶3∶1 之比的变形。产生这样的结果是由于 *A* 基因

和 *B* 基因是同效的，都可以产生一部分棕色色素，同时表达时由于基因的剂量增加，色素加深呈棕红色，这样互为不完全显性，故称为**半上位**（semi-epistatic）或**不完全上位**。或**互加效应**(adititive effect)。

南瓜果形的遗传也属此类。*A*_*B*_为扁形，*aabb* 为长形，*A*_*bb* 或 *aaB*_为球形。*A* 基因和 *B* 基因都是控制果型向横向延伸，突变形是向纵向延伸，*A* 和 *B* 的作用可以互相叠加，所以当 *A*_*B*_ 时果形呈扁形。

P　棕红 × 白色
AABB ↓ *aabb*
F_1　棕红 *AaBb*
F_2　棕红　淡棕　白色
*A*_*B*_　*A*_*bb*　*aabb*
*aaB*_
9 : 6 : 1

图 4-17　半上位效应。

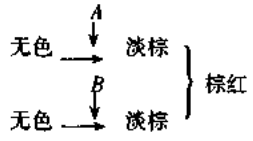


图 4-18　半上位效应的生化机制。

四、等上位

荠的果形有三角形(*AABB*)和筒形(*aabb*)两种，三角形和筒形杂交，F_1 代的果形都是三角形，F_2 代只有两种表型三角形和筒形，比例为 15 : 1(图 4-19)，这也是 9 : 3 : 3 : 1 的变形。从此例可知，除 *aabb* 以外，任何基因形，只要有一个 *A* 或 *B*，表型都为三角形。*A* 和 *B* 的作用相同，本身是完全显性，*A* 和 *B* 互为完全显性，称等上位（isoepistatic）或**双显性上位**（double dominant epistasis），也称**积量效应**或**重复基因**（duplicate genes）。

P　三角形 × 筒形
AABB ↓ *aabb*
F_1　三角形 *AaBb*
↓
F_2　三角形　三角形　三角形　筒形
(*A*_*B*_　*B*_*bb*　*aaB*_)　*aabb*
15 : 1

图 4-19　等上位效应。

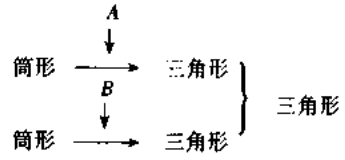


图 4-20　等上位效应的生化机制。

五、隐性上位

将黑色的小鼠和白色小鼠杂交（图 4-21），F_1 代是黑色的，再将 F_1 互交产生的 F_2 有三种表型，黑色，棕色，白化，比例为 9 : 3 : 4。产生的原因是形成黑色素的过程是一个连锁反应。当 *C*、*B* 基因都正常时，才能产生黑色素，当 *B* 基因发生突变而 *C* 正常时虽不能产生黑色素，但有棕色的中间产物形成。当 *C* 发生突变时，即使 *B* 基因正常也因缺乏中间产物而不能合成黑色素，反应停止在无色色素原这一步，所以小鼠呈白色（图 4-22）。若 *C*，*B* 都发生突变，小鼠同样呈白色。在这种类型中当 *C* 基因以隐性等位基因存在时，无论 B 基因是正常还是突变，都不能发挥作用，似乎 *c* 对 *B*，*b* 起到异位显性作用，故称**隐性上位**(recessive epistasis)。

P 黑色×白化

BBCC ↓ *bbcc*

F_1 黑色 *BbCc*

↓

F_2 黑色 棕色 白化 白化

*B_C_* *B_cc* (*bbC_* *bbcc*)

9 : 3 : 4

图 4-21 隐性上位效应。

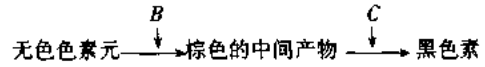


图 4-22 隐性上位效应的生化机制。

六、互补效应

香豌豆有紫花品系(*PPCC*)和白花品系(*ppcc*),杂交的 F_1 代皆为紫花(*PpCc*),自交后 F_2 代,只有紫花 (*P_C_*) 和白花 (*P_cc*, *C_cc*, *ppcc*) 两种表型,比例为 9∶7(图 4-23),当然也是 9∶3∶3∶1 之比的变形。也就是说只有当 *C*, *P* 基因都正常时才开紫花。这是由于紫色素的合成是经过了一系列的连锁反应(图 4-24),无论是 *C* 基因还是 *P* 基因发生突变,此反应都不能完成,而中间产物是白色的,所以只有 *C P_* 基因型呈紫色,其余基因型皆为白色。这种类型称**互补效应**(complement effect)。和隐性上位不同的是其中间产物是没有产生色素(无色),在这种类型中 *C* 和 *P* 两个座位只要其中一个座位以隐性状态存在,对另一个基因都能起到异位显性的作用,故也称为**双隐性上位**(double recessive epistatic)。

P 紫色×白花

CCPP ↓ *ccpp*

F_1 紫 *CcPp*

↓

F_2 紫色 白花

*C_P_* (*C_pp* *ccP_* *ccpp*)

9 : 7

图 4-23 互补效应。

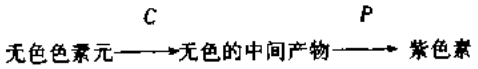


图 4-24 互补效应的生化机制。

七、显性上位

显性上位产生 3 种表型,比例为 12∶3∶1,但产生这种比例的机制可能有不同的途径,一种是抑制基因所引起的。在植物中有一种可用于治疗心脏病的药物叫毛地黄(*Digitalis purpuread*)是一个很好的例子(图 4-25)。毛地黄的 *M* 等位基因控制合成红色花色素苷。*D* 为修饰基因,它控制合成大量的花色素苷,花呈深红;*d* 只能控制合成少量的花色素苷,花呈浅红。*W* 座位控制花色素在花冠上的不同分布。*W* 等位基因使红色色素只分布于筒状花的喉部, 其余部位不分布。*w* 基因使花色素分布于整个花冠。所以 *DdWw* 的表型为白色花喉部带红点,*D_uw* 表现为既有大量的红色素,又能分布于整个花冠,故呈深红;*ddww* 仅有少量红色色素,也能广泛分布,花为浅红色。以上可以看出,只在显性基因 *W* 的存在,无论是 *D* 还是 *d*,都为白色带红点 ,也就是说当 *W* 为显性时对另一个基因呈异位显性,也就称为**显性上位**(dominant epistasis)。

另一种类型的显性上位的分离比也是 12∶3∶1,但机理不同,如南瓜的皮色有白色、黄色、绿色三种,遗传的规律(如图 4-26)。产生的机制是由于 *W* 等位基因是一个

P　　深红 × 白色带红点
(*MM*)*DDww* ↓ (*MM*)*ddWW*
F_1　　白色带红点 *DdWw*
↓互交
F_2　白色带红点　深红　浅红
(*D_W_*　*ddW_*)　*D_dd*　*ddww*
12　:　3　:　1

图 4 - 25　毛地黄的显性上位效应。

抑制基因(inhibitor)作用的结果，原来产生黄色色素要经过两步生化反应(图 4 - 27)当 *W* 基因存在时可以抑制 *G* 基因的表达，结果瓜呈白色。若 *W* 突变为 *w*，则失去了抑制作用，反应可以顺利进行，直到产生黄色的色素，当 *Y* 突变为 *y* 时，则不能形成黄色色素，瓜呈绿色。结果也同样会出现 12 : 3 : 1 的分离比。

还有第三种可能的机制是深色和浅色同时表达，其结果是深色掩盖了浅色。如燕麦颖片颜色的遗传，当黑色颖片和白色颖片的品系杂交 F_1 代为黑颖，F_1 自交后 F_2 产生三种表型，黑颖、黄颖和白颖，比例也为 12 : 3 : 1(图 4 - 29)，产生的原因是(图 4 - 28)黑色掩盖了浅色。*B* 存在时，即使 *Y* 也存在，由于黑色产生掩盖了黄色，表型仍为黑色，只有 *b* 存在时 *Y* 表型才呈黄色，*bbyy* 时既不能产生黑色素，也不能产生黄色色素，所以为白色。

P　　*WWYY* 白色 × 绿色 *wwyy*
WWyy 白色 × 黄色 *wwYY*
↓
F_1　　白色 *WwYy*
↓互交
F_2　白色　白色　黄色　绿色
(*W_P_*　*W_yy*)　*wwY_*　*wwyy*
12　:　3　:　1

图 4 - 26　南瓜皮色的显性上位效应。

W
↓抑制
G　　*Y*
白色色素 ⟶ 绿色色素 ⟶ 黄色色素

图 4 - 27　显性上位抑制基因的生化机制。

B
白色色素 ⟶ 黑色
Y ↓　　黑色
白色色素 ⟶ 黄色
↓

图 4 - 28　燕麦颖片颜色的遗传机制。

P　　黑颖 × 白颖
BBYY ↓ *bbyy*
F_1　　黑颖 *BbYy*
↓
F_2　黑颖　黑颖　黄颖　白颖
(*B_Y_*)　*B_yy*　*bbY_*　*bbyy*
12　:　3　:　1

图 4 - 29　燕麦颖片颜色的遗传。

八、抑制效应

另一种重要的互作是**抑制基因**(suppression gene)。此是一个基因抑制另外一个基因的表达，但其生化途经和显性上位不同。在报春花属(*Primula*)中 *K* 基因可以控制合成一种黄色的锦葵色素，但另一个 *D* 基因存在时可抑制其表达(图 4 - 30)。因此 *KKdd* 为黄色，*KKDD* 为白色，二者杂交的 F_1 代为白色 *KkDd*；F_2 代出现两种表型白色和黄色分离比为 13 : 3(图 4 - 31)。以上八种不同类型的基因互作可以归纳为表 4 - 8。

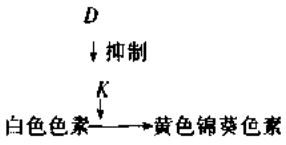


图 4-30 抑制效应的生化机制。

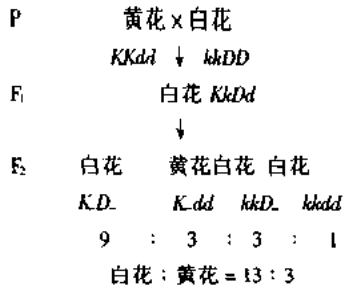


图 4-31 隐性上位效应。

表 4-8 不同等位基因的相互作用

类型	A和B基因的关系	实例	表型比例				生化机理
			A_B_	A_bb	aaB_	aabb	
			9	3	3	1	
共上位(镶嵌上位)	A，B完全显性； A≠B； A和B互为完全显性	鹦鹉羽色 鼠毛色	绿 鼠 9	蓝 黄 3	黄 黑 3	白 棕 1	A 无色素元→蓝色↘ B + →绿色 无色素元→黄色↗ A 棕色→ + 黄斑带→黄色 B↓A 黑色→ + 黄斑带→野鼠色
半上位(互加效应)	A，B完全显性；A≠B； A和B为不完全显性	猪毛颜色 南瓜果形	棕 扁 9	淡棕 球 6		白 长 1	A 无色→淡棕↘ B + →棕红 无色→淡棕↗
等上位(双显性上位)	A，B完全显性；A = B； A和B互为显性	芥果形 燕麦芒 鸡腿毛	三角形 硬芒 有毛 15			筒形 脆芒 无毛 1	A 筒形→三角形↘ B + →三角形 筒形→三角形↗
隐性上位	A，B完全显性；A≠B； a和b为隐性时对A、B起上位作用	兔毛颜色 葱毛颜色 玉米颜色	黑 红 紫 9	棕 黄 红 3	白 白 白 4		B C 无色→棕色→黑色素
互补效应(双隐性上位)	A，B完全显性；A≠B； a和b为隐性时对A，B起上位作用	香豌豆 白三叶草	紫 含氰 9	白 不含氰 7			A B 无色→无色→紫色素

续表

类型	A和B基因的关系	实例	表型比例				生化机理
			A_B_	A_bb	aaB_	aabb	
			9	3	3	1	
显性上位	A、B完全显性；A≠B； A为显性时对B，b起上位作用	毛地黄	白色带红色		深红	浅红	A（抑制） G　B 白色→绿色→黄色色素 A 白色色素→黑色↘ B　　+→黑色 白色色素→黄色↗
		南瓜颜色	白		黄	绿	
		燕麦颖片颜色	黑		黄	白	
			12		3	1	
抑制效应	A、B完全显性；A≠B； A抑制B的表达	藏报春	白		黄	白	A （抑制） B 白色色素→黄色锦葵色素
		蚕茧色	白		黄	白	
		鸡羽色	白		有色	白	
			13		3		

第五章

基因的连锁与交换和真核生物的基因作图

基本概念

在减数分裂时同源染色体之间的交换导致了遗传重组的发生。一个交叉(chiasm)就是交换的位点。交换是同源染色体通过联会、断裂、重接在相关位置彼此交换。

遗传重组的证据是在减数分裂中发现交换时细胞学标记发生重组,遗传标记也发生重组。

交换是一个交互事件,发生在减数分裂前期 Ⅰ染色体形成二价体的阶段。

重组是交换的结果,但交换并不一定引起重组,如两标记间发生偶次数交换,标记并不发生重组。所以一般实际的交换值要大于重组值。在距离较近时,双交换发生的可能性较小,故重组值近似等于交换值。在有双交换发生时,两端标记之间的距离近似等于重组值加二倍的双交换值,重组频率近似于两基因间的交换频率。当基因间距离增加时,多次交换的发生可能使重组频率产生低估以至影响图距比实际的距离要小。所以要计算双交换的频率来加以校正。

两个染色体之间发生一次交叉可以阻止相邻位置第二次交叉的发生,这个现象称为交叉干涉。

第一节 基因连锁的发现

一、相引与相斥

1905 年在遗传学发展的第一个高潮时期,彼特逊(Bateson, W)和庞尼特(Punnet, RC)研究了香豌豆(*Lathyrus doratus*)两对性状的遗传。他们选择的性状,一对是花的颜色,有紫色和红色两种,紫花(*P*)对红花(*p*)是显性;另一对性状是花粉的形状,有长形(*L*)和圆形(*l*)两种,长形对圆形是显性。杂交的结果 F_1 代都是紫色花,长形花粉,证明紫花对红花是显性,长形花粉对圆形花粉是显性的。在 F_2 代中,开紫花的 305 株,开红花的 76 株,带有长形花粉的是 305 株,圆形花粉的为 76 株。

P　　紫长 × 红圆

(*PPLL*) ↓ (*ppll*)

F_1　　紫长

(*PpLl*)

↓⊗

F_2	紫长 (*P*. *L*.)	紫圆 (*P*. *ll*)	红长 (*ppL*.)	红圆 (*ppll*)
观察数:	284	21	21	55
按孟德尔理论数:	215	71	71	24

这个结果花的颜色和花粉形状的分离比各自符合 3∶1,表现都是由单基因控制的,不符合孟德尔的两对因子遗传的分离比 9∶3∶3∶1 之比。庞尼特又扩大样本重做

实验，其结果和前次的相近，仍不符合孟德尔定律，亲组合比理论数多，重组合比理论数少得多，对于这一有趣的现象，庞尼特提出了新的假设，认为似乎两对基因在杂交子代中的组合并不是随机的，而是原来属于同一亲本的两个基因更倾向于进入同一配子中，此叫做**相引**(coupling)，原来属于不同亲本的两个基因(P 和 l，p 和 L)之间在形成配子时相互排斥，称为**相斥**(repulsion)。为了证明这个假设是否正确，他们又将紫花圆形花粉和红花长形花粉的亲本杂交，结果 F_1 代和第一次杂交相同都是紫花长形花粉，F_2 得到四种表型，结果也不符合孟德尔的 9∶3∶3∶1 分离比，虽差异不像前次那样显著，但仍是亲组合比理论数多。也就是说 F_2 代性状的分离比①受到亲本性状的影响；②生殖细胞形成时亲组合的基因相引，重组合的基因相斥。这个解释对于以上的实验能够自圆其说，但不能解释为什么孟德尔研究的两对性状都是自由组合的。其实相斥和相引的理论，没有和遗传的染色体学说相联系，因此也朦胧地道出了连锁的现象，可惜的是尚未真正地揭示基因在染色体上连锁的规律。在科学发现史上类似的情况曾重复过多次，说明一个好的实验是重要的，但对实验的科学分析、合理的推理更为重要，否则一个重要的发现就会和实验者擦肩而过。摩尔根正是对相似的实验根据遗传的染色体学说进行科学的分析，发现染色体连锁这一重大理论，使得遗传学又向前发展了一步。

P	紫圆 × 红长			
	(*PPll*) ↓ (*ppLL*)			
F_1	紫长			
	(*PpLl*)			
	↓⊗			
F_2	紫长	紫圆	红长	红圆
	(*P-L-*)	(*P- ll*)	(*ppL-*)	(*pp ll*)
实得数：	226	95	97	1
按孟德尔理论数：	235.8	78.5	78.5	26.2

二、连锁法则的建立

1910 年摩尔根和他的学生布里吉斯(Bridges CB)研究了两对基因的遗传，发现了连锁和互换，建立了遗传学第三个基本定律，**连锁法则**。摩尔根发现果蝇的红眼和紫眼(purple)、长翅和残翅(vestigial)两对性状都是非伴性性状，各自的遗传都符合孟德尔法则。他们将红眼长翅的果蝇和紫眼残翅的果蝇进行杂交，F_1 代为红眼长翅，然后将 F_1 和双隐性亲本进行测交，所得到的测交后代按孟德尔法则应有四种表型，分离比为 1∶1∶1∶1，但实际得到的结果确有相引和相斥的现象，亲组合多于理论数，重组合少于理论数。同贝特森等一样，为了进一步验证，他们又改变组合重新实验，结果相似，仍是亲组合多于理论数，重组合少于理论数。

P	♀ 红长 × 紫残 ♂			
	($Pr^+Pr^+Vg^+Vg^+$) ↓ ($PrPrVgVg$)			
F_1	红长 × 紫残(测交)			
	(Pr^+PrVg^+Vg) ↓ ($PrPrVgVg$)			
F_2	红长	紫残	红残	紫长
	(Pr^+Vg^+)	($PrVg$)	(Pr^+Vg)	($PrVg^+$)
	1339	1195	151	154

但他们的解释却不相同，摩尔根等认为第一种组合 Pr^+、Vg^+ 两个基因位于同一条染色体上，Pr、Vg 也位于另一亲本的相应染色体上，两个亲本都是纯合体，杂交后 F_1 代的这一

P	♀ 红残 × 紫长 ♂			
	↓			
F_1	红长 × 紫残(测交)			
	↓			
F_2	红长	紫残	红残	紫长
	(Pr^+Vg^+)	($PrVg$)	(Pr^+Vg)	($PrVg^+$)
	157	146	965	1067

对染色体分别携带着红眼长翅和紫眼残翅基因(图 5－1)。

减数分裂时有的细胞在 Pr^+、Vg^+ 两个基因之间发生染色体的交换重组，产生了重组型配子 Pr^+Vg 和 $PrVg^+$(图 5－2)，这种配子的比例比亲组合型的配子少，所以通过测交所得到的后代不是 1∶1∶1∶1，而是亲组合的表型红眼长翅和紫眼残翅为多；

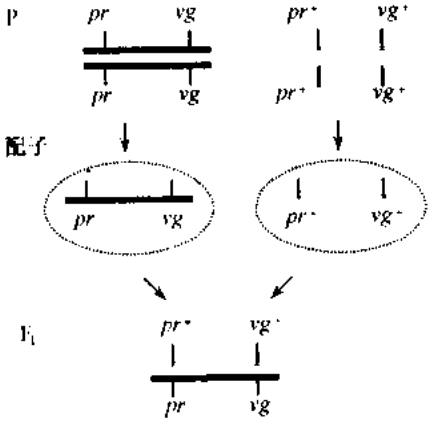


图 5－1　两对位于同一条染色体上的等位基因的简单遗传
(仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 5－2)

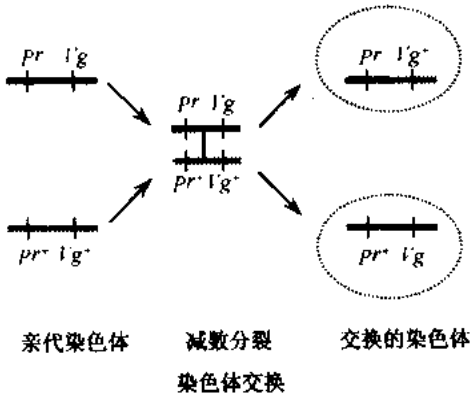


图 5－2　在减数分裂中染色体交换
(仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 5－2)

相反重组合的后代红眼残翅、紫眼长翅为少，另一组合的杂交也是同样的道理。

他们的假设归纳起来主要有三个论点：① 相引是两个基因位于同一条染色体上，相斥则反之。其实按现代的概念，相引就是**顺式**（*cis*），相斥就是**反式**（*trans*）。② 同源染色体在减数分裂时发生**交换**（*crossing-over*）重组（1910 年）；③ 1911 年摩尔根又补充了一点：位置相近的因子相互连锁。连锁和交换的重组称为**染色体内重组**（intrachromosomal recombination），因染色体自由组合而产生的重组称为**染色体间重组**（interchromosomal recombination）。

第二节　基因重组和染色体交换的作用

一、玉米实验

摩尔根的学说是根据实验的结果，假设在减数分裂时由于染色体的交换导致了遗传重组，虽然这一学说已被广泛接受，但毕竟是个假说，还必须进一步地通过实验加以证实。1930 年前都未能获得证实重组和染色体交换之间关系的有力证据。直到 1931 年美国著名的遗传学家麦克林托克（McClintock）指导了她的女博士生克莱顿（Creighton, B）以玉米为材料进行了一项有趣的实验，为染色体交换导致遗传重组提供了第一个有力的证据。

他们已经研究了玉米的第二个最小的染色体，即 9 号染色体上，其带有色素基因 *C* 和糯质基因 *Wx*，在其短臂上（靠近 *C*）带有一个明显的纽结（knob）。在长臂端（靠近 *Wx*）有一条来自第 8 号染色体的附加片段，正常的染色体是没有纽结和易位片段的，因此纽结和附加片段就成为一种细胞学标记。她们选用了一个杂合品系（如图 5－3），其中一条染色体带有有色（*C*）和非糯（*Wx*）基因，两端有标记，

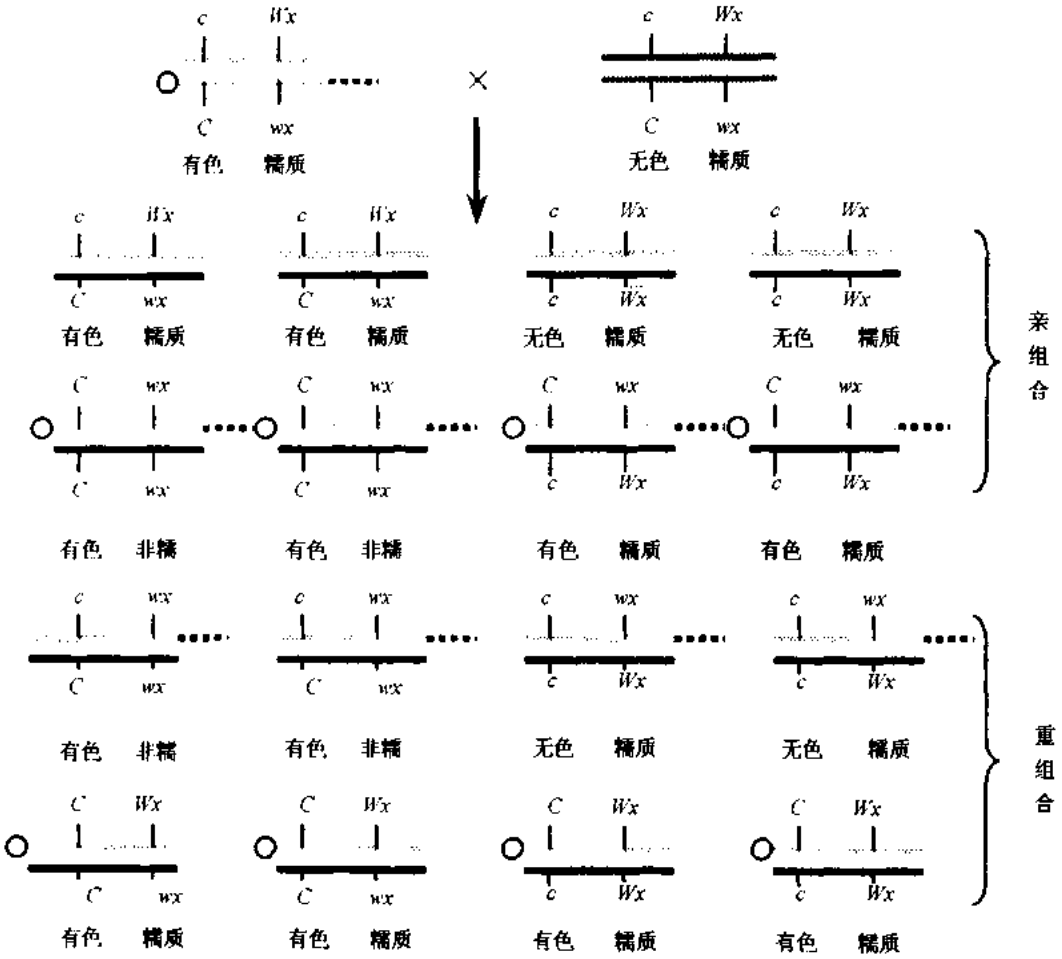
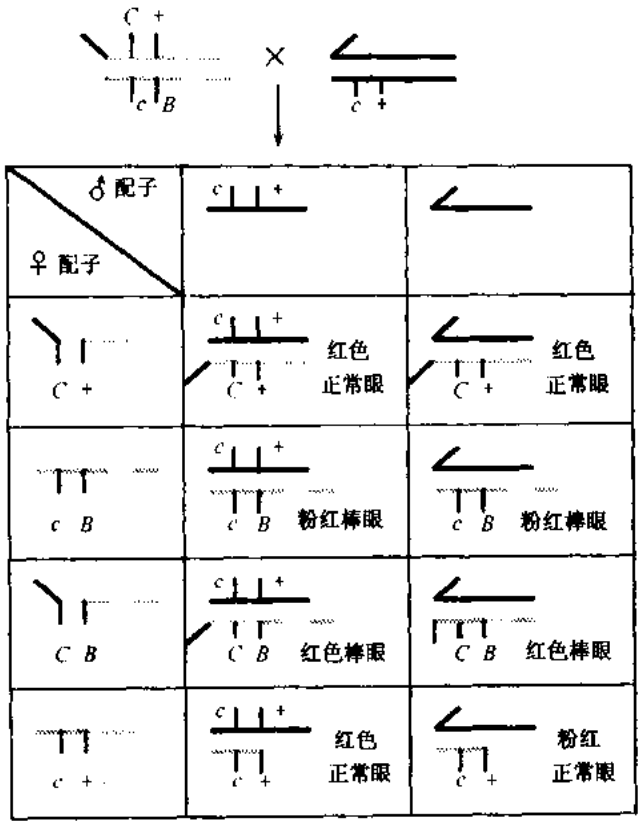


图 5-3　Creighton H B 和 McClintock B 的玉米杂交实验，首次证实了染色体的交换重组（*C*：色素形成基因；*Wx*：糯质基因，因糯质玉米的胚乳呈半透明状，故此基因又称蜡质基因）。

而另一条染色体是正常的，两端不带有标记。这条染色体上带有的是无色基因（*c*）和糯质基因（*Wx*），他们通过杂交后比较亲本型后代和遗传重组后代的染色体，发现亲本型的后代都保持了亲本的染色体排列，而有的重组型后代的染色体也发生了重组。这样他们把遗传学和染色体内重组的细胞学证据联系起来。但"交叉即是染色体交换"这一观点直到 1978 年才获得证据。

二、果蝇实验

在克莱顿和麦克林托克的结果发表没几周，斯特恩（Stern C）又发表了果蝇实验的证据，和玉米实验有异曲同工之妙。斯特恩选择带有异型同源染色体的雌果蝇为材料。此果蝇的一条 X 染色体上带有两个突变基因，一个是 *Car*（Carnation）基因，是隐性突变，纯合体为粉红眼；另一个是 *B*（Bareye）基因，表型为棒状眼，为显性突变。这条 X 染色体缺失了一个片断；另一条 X 染色体的相应等位基因都是野生型的，但染色体带有 Y 染色体的易位片段，这样两条不同的 X 染色体都有了细胞学的标记，可供镜检分辨。雄性果蝇的 X 染色体上带有 *car* 和非棒眼基因，雄性是完全连锁的。在杂交后代中，不仅可以看到表型（眼色和眼形）的重组，同时在显微镜下还可以观察到带有标记



C(carnation)：粉红色眼基因，隐性突变；*B*(Bar)：棒眼基因，显性突变

图 5-4　Stern C. 的果蝇杂交实验，证实了染色体的交换重组。

染色体的重组。这两种重组又完全一致，从而证实了基因重组是交换的结果。

第三节　交换和重组值

染色体内重组是交换的结果，但交换不一定导致重组。因为我们是依据标记基因来判断是否重组的，若在两个标记基因之间发生奇次数交换，结果必然导致重组，但若发生偶次数交换，标记基因并不重组。

不能进行自由组合的基因群则称之为**连锁群**（linkage group），排列在同一条染色体上及其同源染色体上的基因属于同一连锁群。一种二倍体生物的连锁群数应和其染色体对数相等。在少数的生物中如雄果蝇和雌家蚕都不发生重组，人们称之为**完全连锁**。1922 年英国的**霍尔丹**(Haldane JBS)提出：凡是较少发生交换的个体必定是异配性别个体，此就称为**霍尔丹定律**。为什么在雄果蝇和雌家蚕中不会产生重组呢？根据细胞学的观点，发现在减数分裂中没有交叉，也看不到联会复合体。那么在减数分裂中性染色体如何配对又如何精确分离呢？这一机制尚待进一步研究。

根据交换发生的随机性，我们可以说，距离越长，发生交换的机会越多，因此**交换值**(crossing-over value)的大小就可以用来表示基因间距离的长短。但我们无法来直接测定**交换率**，只有通过标记基因的重组来估计交换的频率。

由于有时虽发生了交换（如双交换），但没有导致重组，因此**重组值**(recombination value)并不完全等于交换值，而是重组值≤交换值。重组值或重组率(recombination frequence)P≅d(两个标记基因间距离的估计数)。

$$P = [重组合/(亲组合 + 重组合) \times 100\%$$

单位是图距单位 m.u.(map unit)或厘摩(centimorgan,cM)

重组值和实际图距的关系也反应了重组值和交换值之间的关系。

Haldane 的图谱函数是 $= \frac{1}{2}(1 - e^{-2d})$(图 5-5)

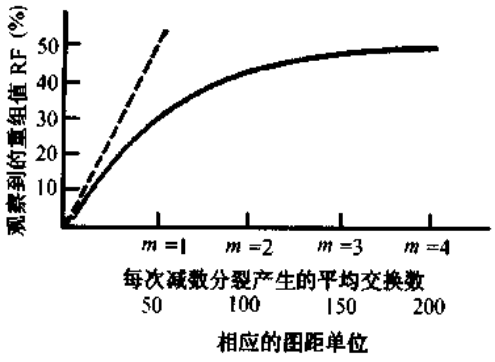


图 5-5　霍尔丹图谱函数。函数曲线表示交换值(RF)和真实图距之间的相关性。(转引自 Griffiths A.J.F.et al.:《An Introduction to Genetic Analysis》,5th ed.,1993,Fig.6-4)

式中 P 是重组频率，d 是图谱距离，e 是自然对数的底（e = 2.71828）。在此函数中，假设没有干涉，这只是对较大的染色体图距是有效的。

从图中我们可以看出按 Heldane 函数作出的曲线，①在图距 10 以内曲线近似直线方程，斜率接近于 1，重组率几乎等于交换率，它可以直接被看作是图距。②当图距大于 10 时，此重组率总是小于交换率，不能直接被看作为图距，必须加双交换的值予以校正。③重组率最大值不可能超过 50%，无论染色体上两点距离如何大，一次交换只可能发生在同源染色体的两条染色单体之间，另外两条单体并未发生交换，即使是每个初级性母细胞都如此，重组率也只有 50%，其实当重组率达到 50%时已是染色体间的自由组合了。

第四节　基因定位和遗传作图

摩尔根的果蝇实验已表明重组是同源染色体发生交换的结果，两个连锁基因之间的距离决定了它们的重组值。那么我们能否用实验的方法来确定真核生物染色体上不同基因的位置呢，答案是肯定的，这种方法我们称之为**遗传作图**（genetic mapping）。

一、三点测交

前面已经介绍基因位于染色体上，而且位置相近的基因是相互连锁的，不难想象基因在染色体上是直线排列的，但如何来证明这个问题呢？摩尔根的学生斯特蒂文特（Sturtevant）想出了一个巧妙的办法，那就是**三点测交**（three-point test cross）。从几何知识上知道，要证明 a、b、c 三点共线，可以通过三点之间的距离加以证明，当 ab + bc = ac 时三点在一条直线上。斯特蒂文特借鉴了这一思路，选择了六个性连锁基因进行果蝇的杂交实验，为了很好地说明问题现在我们以红眼 *v*(vermilion eyes)、翅无横脉 *cv*(cross veinless)和截翅 *ct*(cut, or snipped wing edges) 三个基因为例，杂交及结果如下。

表 5-1 三点测交的结果

表型	实得数	比例	重组发生位点		
			c - cv	cv - ct	v - ct
V cv⁺ ct⁺	580	80.9			
V⁺ cv ct	592				
V cv ct⁺	45	5.9	√	√	√
V⁺ cv⁺ ct	40				
V cv⁺ ct	89	12.6	√		√
V⁺ cv⁺ ct⁺	94				
V cv⁺ ct	3	0.6		√	√
V⁺ cv⁺ ct⁺					
合计	1448	100	18.5	6.5	13.2

根据表中计算的结果，可以绘出 *v*、*cv*、*ct* 的遗传图。但是我们会发现 *v*-*ct*（13.2 m.u.）和 *ct*-*cv*（6.5 m.u.）之和为 19.7 m.u.，大于 *c*-*cv* 的距离（18.5 m.u.），这是什么缘故呢？这是因为 18.5 m.u. 这个数据是 *v*-*cv* 的重组值，即从 *v* 和 *cv* 这两个标记的重组中直接得来的，这两点中如果发生双交换的话，因不会产生重组而无法计算，易被漏掉，但从 *cv*-*ct* 和 *v*-*cv* 的重组中可以看到确实发生了两次双交换（图 5-6），交换率为 0.6 % × 2= 1.2%，如果把漏掉的双交换值加进去则

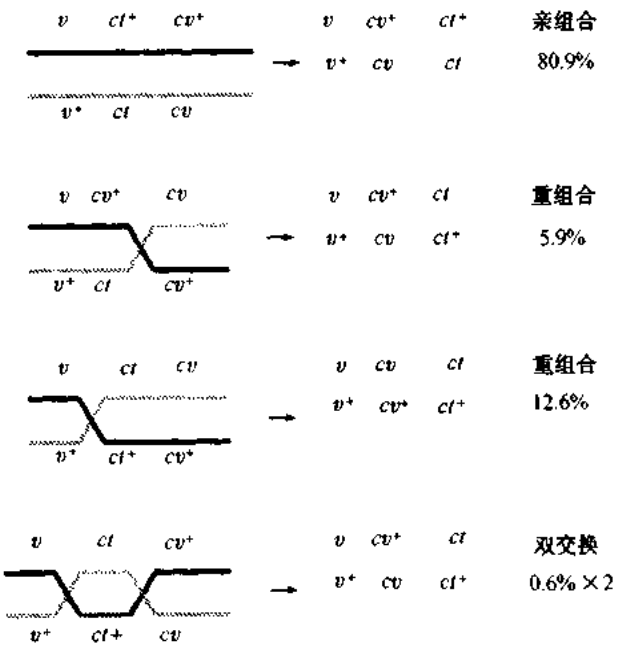


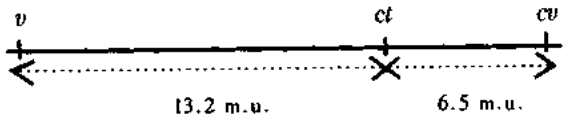
图 5-6 三点测交的图解。

$$(v-cv)=(cv-ct)+(v-cv)$$

18.5 m.u. + 1.2 m.u. = 6.5 m.u. + 13.2 m.u. = 19.7 m.u.

因此我们可用下面的公式来表示 a-b之间的重组值和双交换的关系：

$$ab=ac+cb-2(ac)(cb)$$



ab 是 *a* 和 *b* 两个基因之间的重组值。*c* 是 *a* 与 *b* 之间的一个基因，*ac*、*cb* 表示 *a* 和 *c*、*c* 和 *b* 的重组值，（*ac*）（*cb*）表示双交换值。

三点测交实验的意义在于:①比两点测交方便、准确。一次三点测交相当于3次两点测交实验所获得的结果;②能获得双交换的资料;③证实了基因在染色体上是直线排列的。

二、并发率和干涉

如果在二条非姊妹染色单体上分别发生两次单交换,单交换Ⅰ的概率为α,单交换Ⅱ的概率为β,如果这两次交换同时发生(即为双交换),其概率理论上为(α)(β),假设实际测得双交换的概率为γ。并发系数(coefficidence of coincidence or coincidence)C为观察到的双交换数除以预期双交换这一比值。按上列数据

$C=\gamma/(\alpha)(\beta)$

按上列数据　$C=0.6\%\div[(5.9\%)(12.6\%)]=0.08$

这个结果远小于1,也就是说实际发生的双交换比预期发生的概率要小得多,这是由于一次重组事件的发生,干扰了另一次重组事件的发生,这一现象称为染色体的**干涉**(chromasomal interference)或称为**交叉干涉**(chiasma interference)。以I来表示。干涉和并发系数的关系以下式表示

$$I=1-C$$

当C=1,I=0时,表示无干涉存在

C=0,I=1时,表示存在完全干涉

1>C>0时　表示存在**正干涉**(positive interference),即是一次交换抑制了邻近位点另一次交换的发生。这可能是由于染色单体的断裂和重接物理性地干涉了邻近位点的交换。

C>1,I<0时,表示存在**负干涉**(negative interference),负干涉仅在微生物中发生**基因转变**时(见第23章)才出现。

三、染色单体干涉

在减数分裂的前期,二价体是有4条染色单体,交换并不限于二条非姊妹染色单体之间,而有各种不同的交换形式(图5-7)。为了便于说明,我们巴4条染色单体分别标上序号。当2-3之间已发生了一次交换的基础上,再发生2-3二线双交换;1-4四线双交换;1-3三线双交换和2-4三线双交换。它们的机会是均等的。一般说来第一次交换仅对于2-3二线双交换有干涉,对其他类型的双交换没有干涉。

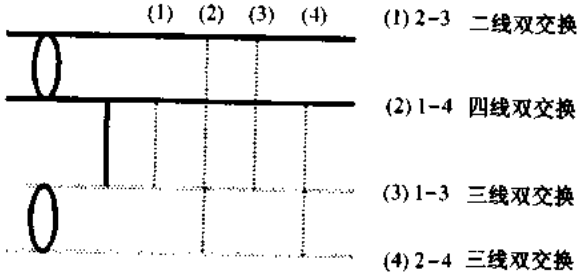


图5-7　当第一次交换发生后,第二次交换可在任意两条非姊妹染色单体之间进行。

第五节 真菌的遗传分析

一、链孢霉（*Neurospora crassa*）的生活史

链孢霉有两种繁殖方式，一种是无性繁殖，当其孢子（N）或菌丝落在营养物上，孢子萌发，菌丝生长形成菌丝体(N)。另一种是有性繁殖，两个亲本必须是不同的**交配型**(matig type)A 和 a，各自的分生孢子会散落在不同交配型**子囊果**(perithecium，**子实体**)的受精丝上，进入子实体，进行核融合，形成 2n 核，(A/a)。二倍体时期十分短暂，很快进行减数分裂，最后再经过一次有丝分裂，在**子囊**(ascus)中产生 8 个单倍体的**子**

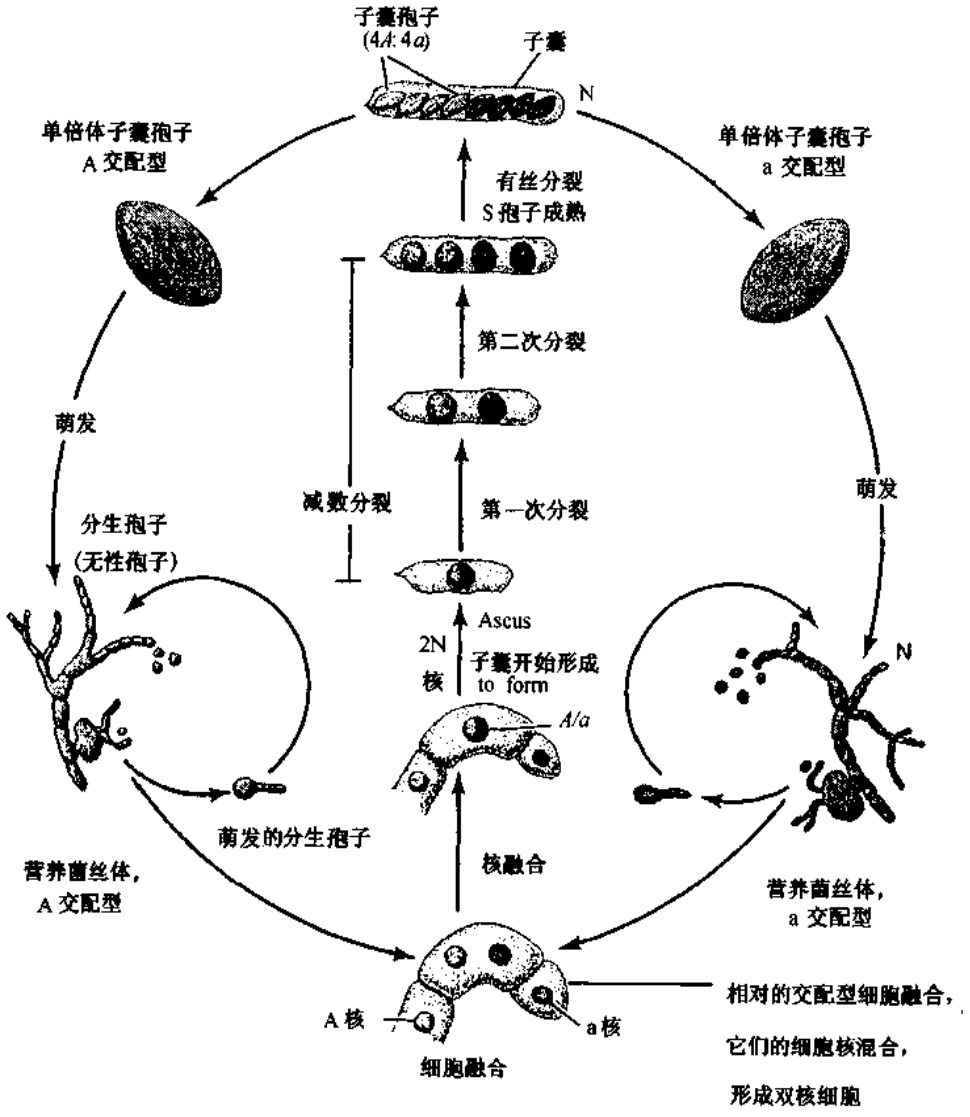


图 5－8 链孢霉的生活周期。(引自 Russell, P. J.; Genetics, 5th ed., 1992, Fig. 5－9)

囊孢子，子囊孢子成熟后又可萌发，长成新的菌丝体(图 5-8)。

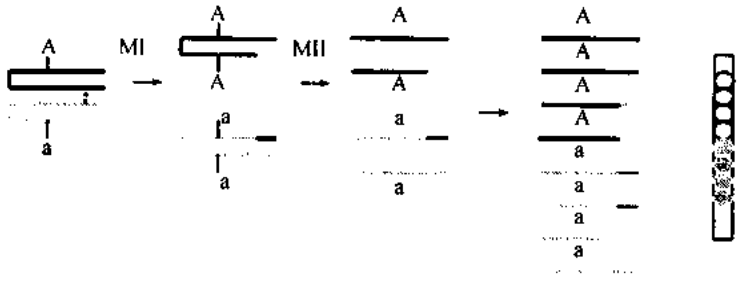
链孢霉后代多，样本大，统计分析的误差小，生活周期段短，在短时间内可获得结果。链孢霉易培养和操作，可利用选择培养基筛选各种突变型；链孢霉又属真核生物，可作为真核的研究模型，因此是很好的遗传学实验材料。子囊孢子是单倍体，不存在显隐性的问题，表型和基因型一致；孢子按顺序排列在子囊中，我们称其为**顺序四分子**(ordered tetrad)，经过分析易于确定是否是重组类型及基因的转变，其着丝粒本身也可看作是一个位点来研究，因此在四分体时期进行遗传分析是很方便的，这种方法称为**四分子分析**(terard analysis)。

二、着丝粒作图

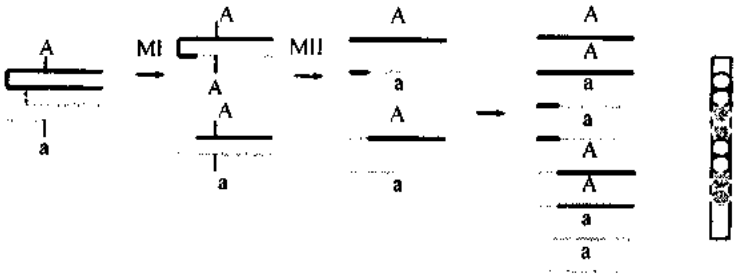
链孢霉的野生型又称为**原养型**(prototroph)，在基本培养基上就能生长，子囊孢子按时成熟呈黑色。有一类突变型是不能合成某些物质，称为**营养缺陷型**(auxotroph)，在基本培养基上不能生长。如赖氨酸缺陷型是自己不能合成赖氨酸，因此培养时必须在基本培养基上加入适量的赖氨酸，此种缺陷型才能生长。缺陷型比野生型的子囊孢子成熟得慢，所以镜检时突变型的子囊孢子呈白色，和黑色的野生型子囊孢子很容易分辨。

1. 第一次分裂分离

在同源染色体上如果有一对等位基因 A 和 *a*，如果交换发生在 *A* 和着丝点之外，那么在第一次减数分裂时 *A* 和 *a* 就完全分开进入不同的两极，至于各向哪一极移动是随机的，但一旦确定就决定了以后子囊孢子排列的位置。第二次减数分裂时，染色单体各自分开，分别进入 4 个孢子，进一步有丝分裂形成了 8 个子囊孢子，野生型为黑色，营养缺陷型为白色，4 黑 4 白有序排列(图 5-9)，这种类型表明在 *A* 和着丝点之



(a) 第一次分裂分离



(b) 第二次分裂分离

图 5-9　红色链孢霉的第一次分裂分离和第一次分裂分离。

间未发生交换，或没有有效的交换。此称为**第一次分裂分离**(first division segregation)。

2. 第二次分裂分离

和上面的情况相反，若在 *A* 基因和着丝点之间发生了一次交换，那么第一次减数分裂时，一对同源染色体的两条姊妹染色单体携带了 *A* 和 *a*，进入同一个子细胞，*A* 和 *a* 没有分离，同样它们趋向两极时是随机的，直到第二次减数分裂形成四分子时，着丝粒也分开，*A* 和 *a* 才分开，各自进入一个子囊孢子中，故称为**第二次分裂分离**(second division segregation)。第二次分裂时 *A* 和 *a* 趋向哪一级也是随机的，一旦确定就决定了最后的 8 个子囊孢子的排列方式。由于第一次和第二次减数分裂的随机性，最后 8 个子囊孢子的排列有几种不同的组合（图 5-10），但决不会是 4 黑 4 白，总是两两相间。第二次分离表明在标记基因和着丝粒之间发生了一次交换。

第一次分裂分离：+ + — —　　第二次分裂分离：+ + — —

\+ — — +

— + + —

— + — +

图 5-10　第二次分裂分离产生的 8 个子囊孢子有 4 种不同的排列类型。

3. 着丝粒作图(centromere mapping)

就是利用分裂分离来确定是否重组，再来计算标记基因到着丝点之间的图距。计算公式为：　重组率 = (交换型子囊数 / 总子囊数) $\times \frac{1}{2} \times 100\%$

这一公式和真核生物的重组率的计算原理相同，都是计算重组型占后代总数的比。本公式为什么乘$\frac{1}{2}$呢？因为我们统计的单位不是子囊孢子，而是子囊。一个子囊中含有 8 个子囊孢子，它们来自四条染色单体，即使是重组型的子囊里面含的四条染色单体中也只有两条发生了交换，还有两条未发生交换，所以必须乘以$\frac{1}{2}$。

表 5-2　*nic* + × + *ad* 得到不同子囊型的后代

	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
四分体	+	*ad*	+	+	+	+	+	*ad*	+	*ad*	+	+	+	+
基因型	+	*ad*	+	+	+	*ad*	*nic*	*ad*	*nic*	+	*nic*	*ad*	*nic*	*ad*
	nic	+	*nic*	*ad*	*nic*	+	+	+	+	*ad*	+	+	+	*ad*
	nic	+	*nic*	*ad*	*nic*	*ad*	*nic*	+	*nic*	+	*nic*	*ad*	*nic*	+
分裂分离	M_I	M_I	M_I	M_I	M_I	M_{II}	M_{II}	M_I	M_{II}	M_{II}	M_{II}	M_{II}	M_{II}	M_{II}
子囊类型	(PD)		(NPD)		(T)		(T)		(PD)		(NPD)		(T)	
子囊数	808		11		90		5		90		1		5	
染色体交换														

三、链孢霉的连锁作图

链孢霉中有两种缺陷型，一种是 *nic*(nicotinic)，不能合成烟酸；另一种是 *ad*(adenine)，不能合成腺嘌呤，培养时必须在培养基中分别加入烟酸和腺嘌呤。将这两个品系杂交，杂交后代中得到 7 种不同的子囊类型(表 5-2)。**PD** 为**亲二型**(parental ditype)，意为表型像亲本，共有 + *ad* 和 *nic* + 两种类型；**NPD** 为**非亲二型**，意为表型都不像亲本，有 + + 和 *nic ad* 两种类型；T 为四型(tetratyne)，意为有四种基因型 + +，*nic ad*，+ *ad*，*nic* +。

事先我们并不知道 *nic* 和 *ad* 基因是连锁的，通过四分体分析我们不仅能确定这两个基因是否连锁，而且如果连锁的话我们还能确定它们在染色体上的排列顺序，计算出 *nic* 和 *ad* 与着丝点之间的图距以及 *nic* 和 *ad* 间的距离。

首先我们来看第一种子囊型，从 *nic* 和 *ad* 两个基因的组合(横向排列)来看，它们是 + *ad* 和 *nic* + 两种类型，和两个亲本完全相同，故标以亲本二型(**PD**)，这组亲组合的子囊数为 808 个，占了整个后代子囊数(1000)的 80.8%，即大部分为亲组合，少部分为重组合。由此可以判断 *nic* 和 *ad* 两个标记是连锁在同一条染色体上。我们再纵向来看第一纵行，为 + +　*nic nic*，这种排列方式是符合第一次分裂分离(M I)，*nic* 和着丝粒之间未发生交换重组。第二纵行为 *ad ad*　+ + 也属于第一次分裂分离的排列方式，也表明 *ad* 与着丝粒之间未发生交换重组。

第二种子囊型 *nic* 和 *ad* 两个基因的组合方式(横向排列)也只有两种类型，+ + 和 *nic ad*，和亲本都不同，故称为非亲二型(NPD)。从纵向来看，第一行的排列方式为 + +　*nic nic*，第二行为　+ +　*ad ad*，都属于 M I，即两个标记 *nic*、*ad* 与着丝粒之间都未发生交换重组。既然第一类和第二类子囊都未发生重组，那么为什么它们两个标记基因的组合(横向排列)又完全不同呢？这是由于在 *nic* 座位和 *ad* 座位之间发生了 1，4 四线双交换(见表中的图)的结果，使得 *ad* 与着丝粒之间仍然为第一次分裂分离，实际上是发生了双交换，只是着丝粒与 *ad* 座位未发生重组。

第三类子囊的两个标记有四种不同的组合方式，成为四型(T)，其中第一纵行为 + +　*nic nic*，属于 M I，表明 *nic* 与着丝粒之间未发生交换重组。而第二纵行就不同了，排列为 + *ad* + *ad*　，属于第二次分裂分离(M II)，表明在着丝粒和 *ad* 之间曾发生过一次交换(1，4 二线单交换)，才能从第一类亲本的类型变为第三类子囊型。

第四类子囊也属于四型(T)，但和第三类子囊不同，其第一纵行的排列为 + *nic* + *nic*，属第二次分裂分离，第二纵行为 *ad ad*　+　+　属第一次分裂分离，也就是说 *nic* 和 *ad* 之间也发生了一次交换，但对于 *ad* 和着丝粒之间来说是发生 3，4 双线双交换，所以虽然发生了交换，但 *ad* 标记未发生重组。

第五类子囊两个标记的组合只有两种类型，+ *ad* 和 *nic* +，和亲本的组合相同，所以称为 PD，但第一纵行和第二纵行的排列都属于 MII，即 *ad*，*nic* 与着丝粒之间都发生了一次交换，使这两个座位都发生了重组，实际上是在 *nic* 和着丝粒之间发生了一次 2，3 二线单交换，使 *nic* 和着丝粒发生了重组，而 *ad* 和 *nic* 是连锁的。因此也同时发生了重组。

第六类子囊型中，两个标记的组合也只有两种 + + 和 *nic ad*，和亲本的都不同，故属于 NPD，第一纵行和第二纵行的排列都是属于 MII，即 *nic* 和 *ad* 两个座位与着丝点

之间都发生了重组，这种类型的产生较为复杂，从两个标记的组合和纵向的排列来看，是在第四类子囊型交换重组的基础上增加了一次 1,4 四线双交换(表中的图示)，这次双交换发生在 *nic* 和 *ad* 之间。

第七类子囊两个标记有四种组合，即为 T，两列纵向排列分别为 + *nic* + *nic* 和 + *ad* *ad* + ，都属于 M II，即两个标记和着丝粒之间都发生了交换重组。根据两个标记的组合及纵向排列方式不难推测是在第五类子囊型交换重组的基础上产生了 1，3 三线双交换，这次交换应发生在 *ad* 与着丝粒之间。

根据着丝粒作图公式我们可以计算出各标记与着丝粒之间的图距。着丝粒与 *nic* 座位之间的图距可通过各种子囊型中重组型的子囊数来计算，根据表 5-2 只有第 4，5,6,7 型子囊在 *nic* 和着丝粒之间才存在 MII，即发生交换重组，所以

$$\begin{aligned}\text{RF(着丝粒} - nic) &= [(4)(5)(6)(7)/1000] \times 1/2 \times 100\% \\ &= [(5+90+1+5)/1000] \times 1/2 \\ &= 5.05\% = 5.05\ (\text{m.u.}) \\ \text{同理 RF(着丝粒} - ad) &= [(3)(5)(6)(7)/1000] \times 1/2 \times 100\% \\ &= [(90+90+1+5)/1000] \times 1/2 \\ &= 9.30\% = 9.30\ (\text{m.u.})\end{aligned}$$

计算的结果告诉我们 *nic* 和着丝点之间的距离为 5.05 图距单位(m.u.)。*ad* 和着丝点之间的距离为 9.03 图距单位(m.u.)。目前知道了 *nic* 和 *ad* 是在同一条染色体上以及它们和着丝点的距离，但还不知道具体的顺序排列。因根据目前的信息有两种可能：①*nic* 和 *ad* 分别位于着丝点的两侧；②*nic* 和 *ad* 位于着丝粒的同侧(图 5-11)。究竟属于那一种情况呢？假设两个位点在着丝点的两侧，那么 *nic* 和着丝点之间的重组

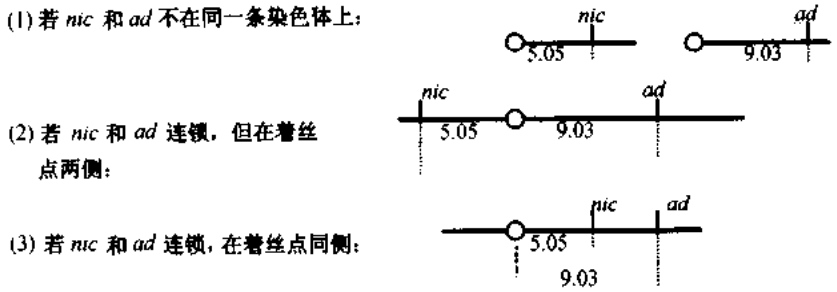


图 5-11 *nic*、*ad* 座位着丝点作图分析。

与 *ad* 和着丝点之间的重组是各自独立的，也就是说 *nic* 和着丝点之间发生一次重组不会影响到 *ad* 和着丝点之间的关系；相反，若两个座位都在着丝点的同侧，一旦 *nic* 和着丝点之间发生了一次交换，若不存在双交换的话，势必使 *ad* 和着丝粒也产生重组。从前面的资料来看，*nic* 和着丝粒之间产生重组的子囊为(4)(5)(6)(7)，共 101 个子囊，其中(5)(6)(7)型子囊共 96 个同时也发生了 *ad* 和着丝粒之间的重组，表明基本是符合第二种情况，那么为什么有五次不同步发生重组呢？从表 5-2 中不难看出，不同步的仅为第四种子囊型，两个座位分别为 MII、MI，共 5 个子囊，从重组的图中很清楚地看到是由于在 *nic* 和 *ad* 之间发生了一次双交换，结果使 *ad* 和着丝粒之间未能重组。

因 *ad* 和着丝粒图距为 9.03m.u.，*nic* 和着丝粒的图距为 5.05m.u.，几乎相差一

倍，而重组是一个随机的过程，因此 *ad* 与着丝粒之间发生重组的交换点可能在着丝点 – *nic* 的区域中也可以发生在 *nic* – *ad* 的范围内。若是前者，则 *nic* 和 *ad* 同时发生重组。*ad* 和着丝粒重组的子囊(MII)共 186 个，其中 *nic* 未同步重组的(MI MII)为 90 个子囊，近似于 *ad* 重组子囊的一半，这和两个座位的距离之比是吻合的。

下面我们再来计算 *nic* 和 *ad* 之间的距离。根据计算重组值的原则：

重组值 = 重组合/(亲组合 + 重组合) × 100%

那么 *ad* 和 *nic* 之间的重组依据是什么呢？表 5 – 2 已表明两个标记的亲组合只有两种类型 + *ad* 和 *nic* + ，不同于这两种类型的子囊都是重组型。因此用下面公式可计算 *nic* 和 *ad* 之间的图距：

$$RF = (NDD + 1/2T)/\text{总子囊数} = \{[(2) + (6)] + 0.5[(3)(4)(7)]\}/1000$$
$$= \{(1+1) + 0.5(90+5+5)\}/1000$$
$$= 5.2\% = 5.2\ (\text{m.u.})$$

NPD 的子囊(2 和 6 型)4 条染色单体都是重组型的，即整个子囊都是重组型，故无需再乘以 0.5，而 T 子囊中只有一半的染色单体为重组型，故需乘以 0.5。

从公式计算出 *nic* 和 *ad* 之间的图距为 5.20m.u.，而从图 5 – 10，我们用 *ad* – 着丝点之间的图距减去 *nic* – 着丝粒之间的图距也应等于 *nic* – *ad* 的图距，即为 9.03 – 5.05 = 3.98(m.u.)，比用公式计算出的图距要小，或者说着丝粒 – *ad* 的图距应为 5.05m.u. + 5.20m.u. = 10.25m.u.，这样就大于前面计算的结果 (9.30m.u.)。这是为什么呢？这是由于双交换的存在使着丝粒 – *ad* 之间的重组值低估了，从表 5 – 3 中可以看出低估的原由。

表 5 – 3　着丝粒 – *ad* 间重组值的计算与低估

子囊型	重组的染色单体数/子囊			遗漏数	子囊数	总的重组型染色单体数			总遗漏数
						· – *nic*	*nic* – *ad*	· – *ad*	
2	0	4	0	4	1	0	4	0	4
3	0	2	2	0	90	0	180	180	0
4	2	2	0	4	5	10	10	0	20
5	2	0	2	0	90	180	0	180	0
6	2	4	2	4	1	2	4	2	4
7	2	2	2	2	5	10	10	10	10
总计	8	10	8	10	192	202	208	372	38

从表中可以看出着丝粒 – *nic* + *nic* – *ad* = 202 + 208 = 410≠着丝粒 – *ad*(372)，这是因为遗漏了 38 条重组型单体。现在的计算方法不是以子囊数而是染色单体数，所以分母应用子囊数乘 4，即低估的重组值为 38/4 × 1000 = 0.95，9.30 + 0.95 = 10.25m.u.。

四、无序四分子分析

无序四分子或八分子中孢子没有特殊的顺序，因此也就不能进行着丝粒作图，但无序四分子也可用于各种减数分裂分离和重组分析。现在我们假设酵母有两个基因位点 a 和 b，我们想确定它们是否连锁，如果连锁还要计算 ab 之间的图距。酵母产生的是无序四分子，怎样才能用这种四分子来研究连锁呢？首先我们来看看在交换中有可能产生那几种不同类型的无序四分子，在双杂种中是否有锁。

无序四分子只可能有三种类型(图 5－11)：**PD**、**NPD** 和 **T** 型，而且子囊里的孢子是无序的。第一种类型 PD 看起来好像是两个位点的第一次分裂分离，其实不然。我们可以不考虑其顺序，这些子囊完全是根据它们含有的二种或四种基因型来分类。在后代中二型性 (D) 基因型可能和亲本相同 (PD) 也可能不同 (NPD)，总的类型不会超过 PD、NPD 和 T 三种。当两个基因 a、b 在不同染色体上时，PD 和 NPD 的四分体不是因为交换而产生，而是由于在减数分裂中期染色体的不同排列所造成的(图 5－12)。

由于四条染色体在中期板上排列的状态是独立的，一对同源染色体分别向两极的移动是随机的，所以 PD 和 NPD 型的四分子的频率相等，NPD 应占 50%，即重组类型占 50%。如果两个基因是连锁的那么结果就不同了。如单交换发生在基因和着丝粒之间，产生四型(T)四分体其频率是有基因和着丝粒之间的距离而定(图 5－13)。

如果两个基因连锁，各种四分体的类型如图 5－13 所示，若未发生交换，产生亲二型(PD)，如在两个基因间发生一次交换，结果产生两个亲组合和两个重组合的染色体称为四型(T)。若发生双交换的话，仍是亲二型(PD)；若发生三链交换的话，则涉及到四条染色体中的三条，有两种形式：1－3 三链双交换和 2－4 三链双交换，其结果是产生四型的子囊，即两条染色体是重组的，两条染色体是亲组合的。最后一种是四链双交换，产生的子囊是 NPD。NPD 是 4 种双交换中的一种，应当比较少。如果 PD 的频率远远大于 NPD 子囊的频率，即 PD≫ NPD，我们就可以推断这两个基因是连锁的。

一旦已确定两个基因是连锁的，只要具有各种类型子囊数据的资料，就可以根据作图公式来计算两个基因之间的距离：(重组体的数目/后代的总数) × 100%。

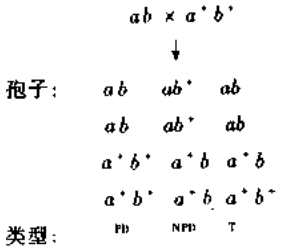


图 5－12　无序四分子三种可能的类型。

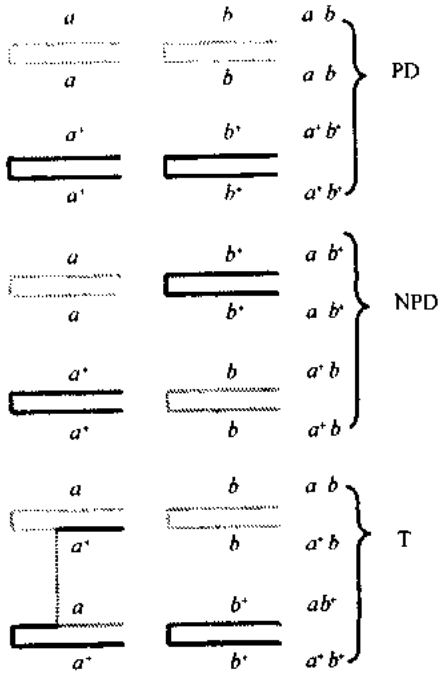


图 5－13　在减数分裂中，中期染色体的不同排列形成三种不同类型的四分体。

现在来看，PD 型子囊四条染色单体都是亲组合，而 T 型子囊有两条染色单体是亲

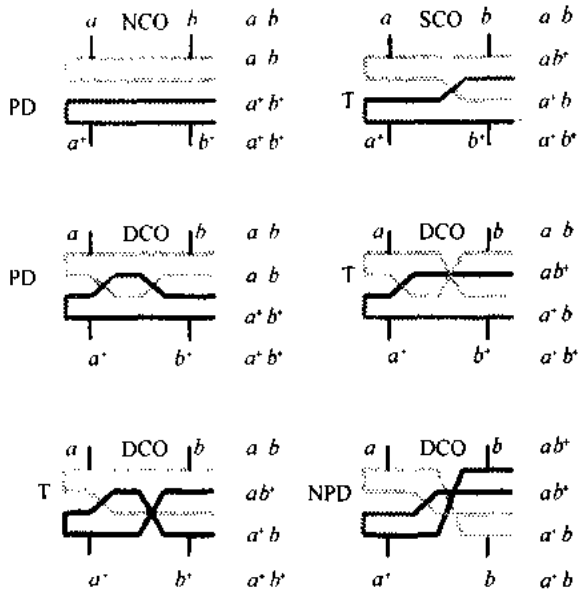


图 5－14　若在两个连锁基因之间发生单交换和双交换，则可产生几种不同类型的四分体 NCO：未交换、SCO：单交换；NCO：双交换。（参考 Griffiths A.J.F. *et al.*：An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 6－14）

组合，另两条染色单体是重组合；NPD 型子囊四条染色单体都是重组合的，那么两个基因之间的重组频率是：(1/2 T + PD)/总子囊数 × 100%。

例如有 200 个子囊，140 个为 PD 型，48 个为 T 型，12 个为 NPD 型，那么两个基因间的重组频率为是：RF = {[1/2(48) + 12]/200} × 100% = 18% = 18 m.u.。

这样计算的结果是两个基因间的重组值，并不能反映真实的距离。真实的距离必须考虑到双交换的存在，可以按以下方法分析推算。前面阐述的重组公式中已经表明两个基因间的重组率为 $ab = ac + bc - 2(ac)(bc)$，也就是说 $a - b$ 间的真实距离应为 a b 之间的重组值再加上 2 次 a b 之间的双交换值 $(ac)(bc)$，或者用 $a - c$ 之间的单交换值加 $b - c$ 之间的单交换值，即

$(a - b)$距离 $= ab + 2(ac)(bc) = ac + bc$

$(a - b)\text{SCO} + 2(a - b)\text{DCO} = (a - c)\text{SCO} + (b - c)\text{SCO}$

SCO：单交换(single crossover)，DCO：双交换(double crossover)

资料给我们的仅两个位点，因此无法用 $(a - c)$ SCO + $(b - c)$ SCO 来计算，而只能从 $(a - b)$SCO + $2(a - b)$DCO 来着手分析，

即　　SCO + 2DCO　　(1)

无序四分子单交换的产物是 T，但另外两次三线双交换也是产生 T 型子囊，我们无法予以分辨，但我们可以这样来推导：产生 T 型子囊的单交换类型等于总的 T 型子囊数减去两种双交换的类型，即

$$\text{SCO} = \text{T} - 2\text{DCO} \quad (2)$$

而一次　　$$\text{DCO} = \text{NPD} \quad (3)$$

但双交换共有四种，那么产生 T 型子囊的双交换又如何求得呢？我们可以用 NPD 来推算，因 NPD 既是双交换的结果又是一种特殊的子囊型，容易区分出来用以计算频率。六种类型中一共有四种双交换，而 NPD 又占了双交换的 1/4，因此整个双交换的频

率是：

总 $$DCO = 4NPD \tag{4}$$

现将(2)、(3)代入(1)式得：

$$SCO + 2DCO = (T - 2DCO) + 2(4NPD)$$
$$= (T - 2NPD) + 2(4NPD)$$
$$= T + 6NPD$$

得到的结果是在此区域内的交换数 m，根据霍尔丹方程（图 5-5）我们可以将 m 换算为图距单位。

因为 $$RF = 1/2(1 - e^{-m})$$

当 m = 0.05 时 $$e^{-m} = 0.95$$

$$RF = \frac{1}{2}(1 - 0.95) = \frac{1}{2} \times (0.05) = \frac{1}{2}m$$

当 m = 0.10 时 $$e^{-m} = 0.90$$

$$RF = \frac{1}{2}(1 - 0.90) = \frac{1}{2} \times (0.10) = 0.05 = \frac{1}{2}m$$

m 是交换值(霍尔丹的方程的横坐标)；m.u. 是重组值(霍尔丹方程的纵坐标)，当 m = 1 时 m.u. = 50%，即 50m.u. 单位。假设在 $ab \times a^{+}b^{+}$ 杂交中，获得的 PD 型子囊为 56%，T 型子囊为 41%，NPD 型为 3%，则

$$RF_{ab} = 50\{0.41 + (6 \times 0.03)\} = 29.5 \text{ m.u.}$$

再让我们来用这一值和直接从 RF 中获得的值进行比较，按公式

$$RF = \frac{1}{2}T + NPD = 0.205 + 0.03 = 0.235 = 23.5 \text{ m.u.}$$

这个结果比上式少了 6m.u.，这是因为在 RF 分析中漏掉了双交换。

五、有丝分裂分离和重组

1. 有丝分裂重组的发现

在正常情况下分裂分离和重组都是在减数分裂中发生的，但实验证据表明在有的有机体中交换也可发生在有丝分裂中。第一例有丝分裂交换的例子是由 Curt Stern 在 1936 年发现的，在交换中果蝇的 X 连锁隐性突变基因黄体色 *y* 取代了正常的灰体色基因，短的曲形刚毛(*sn*)取代了长的弯的刚毛。野生型灰体果蝇的刚毛都是黑色的，而黄体突变果蝇的刚毛则是黄色的。

将一个灰体曲刚毛($y^{+}sn/y^{+}sn$)雌果蝇和一个黄体直刚毛(ysn^{+}/Y)雄果蝇杂交，Stern 发现了正如所预料的那样雌性后代大部分为野生型，但某些雌果蝇的身体有一小块区域是黄色或曲刚毛。这个现象被解释为可能是染色体不分离或或染色体的丢失所致(这些改变是属体细胞的而不是生殖细胞的。)

另一些雌果蝇有**孪生斑**(twin spots)，即两个相连的区域，一个为黄色直刚毛，另一个为黑色焦刚毛，呈现镶嵌表型，在孪生斑的周围都是野生型表型(图 5-15)。Stern 注意到一个有趣的现象，即孪生斑的两部分是相连的，那么这种孪生斑一定是某种遗传事件的交互产物。最好的解释是通过有丝分裂交换而产生，Stern 熟知在果蝇胚胎分化

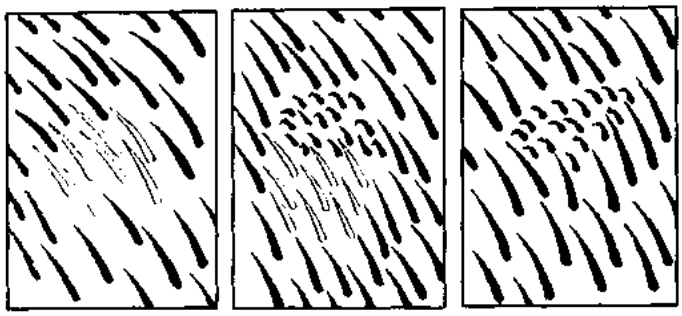


图 5－15　在果蝇中由于有丝分裂交换所产生的孪生斑(a)单个的黄色斑;(b)孪生斑;(c) 单个的曲刚毛斑。(引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 6－18)

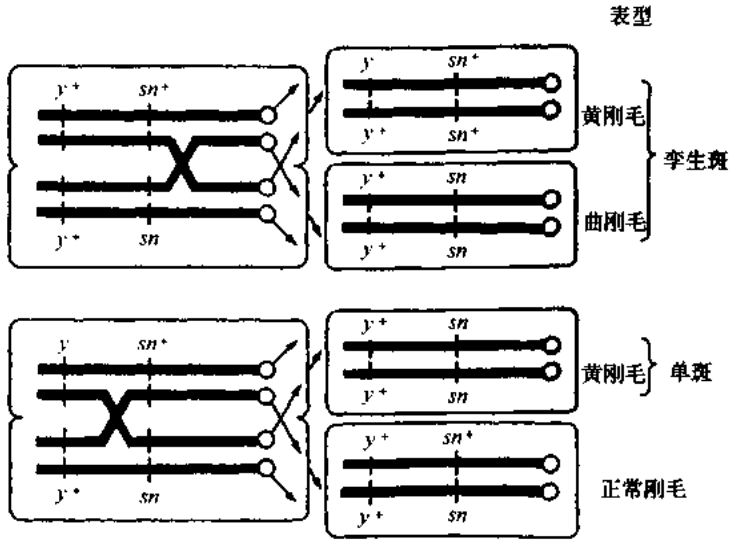


图 5－16　孪生斑产生的机制。(引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 6－19)

之后通过有丝分裂产生各种组织,所以他才提出这一解释。

通过有丝分裂交换产生孪生斑如图 5－16 所示。若有丝分裂发生在着丝点与 *sn* 位点之间,那么产生的两个子细胞的基因型将不同,它们进一步分离形成两个相近的克隆时,表现出孪生斑的特点。若交换发生在 *sn* 和 *y* 之间就会产生单个的黄斑。

在很少的情况下,在染色体复制后,但在中期前,同源染色体形成四分体,进行交换,然后再进入中期,复制后的两条染色单体各趋向一极分离,形成了有丝分裂交换。

2. 真菌系统中的有丝分裂重组

(1)构巢(*Aspergillus nidulans*)曲霉菌的生活周期

有丝分裂重组研究得较为深入的是真菌。我们将用**曲霉属**(*Aspergillus*)为例来说明。曲霉的有丝分裂分析是非常适合的。它的**气生菌丝**产生长链状的**分生孢子**(conidia),这是一种**无性孢子**。每一个分生孢子都有一个单个的核。任何一个孢子的表

型都和它的基因型是一致的，这样便于鉴别。若将两个单倍体品系混合，菌丝融合，两种类型的核都同时存在胞质中，这种品系称为**异核体**(heteokaryon)。和其他品系一样，异核体将产生单核的分生孢子。在少数情况下异核体中的两个单倍体核融合产生一个二倍体核，这一过程称为**二倍化**(diploidization)。二倍体品系的曲霉产生的无性孢子将具有二倍体核。二倍体细胞是可能要经历有丝分裂交换，产生遗传重组。因为二倍体的核不稳定，最终通过有丝分裂产生单倍体核，此单倍体核含有来自两个单倍体亲本的等位基因的重组。由二倍体核形成单倍体核的过程称为**单倍化**(hapliodization)。

在单倍化中，两个同源染色体自由化时，类似减数分裂，然后随机分离，各趋向一极。这种不同于减数分裂而导致基因重组的系统称为**准性生殖系统**（parasexual systems)。真菌的准性生殖周期由以下步骤组成：①在多核菌丝中形成异核体，②在异核体中不同基因型的单倍体核偶尔融合，③和单倍体核并排的二倍体融合核中进行有丝分裂交换，④二倍体融合核不经过减数分裂而完成单倍化。

(2)构巢曲霉的有丝分裂分析

进行有丝分裂重组分析时需要构建杂合子供研究。在构巢曲霉中，异核体由两个不同品系构成，然后再筛选出二倍体核。

如下面的两个单倍体品系

品系 1：　w　ad^+　pro　$paba^+$　y^+　bi

品系 2：　w^+　ad　pro^+　$paba$　y　bi^+

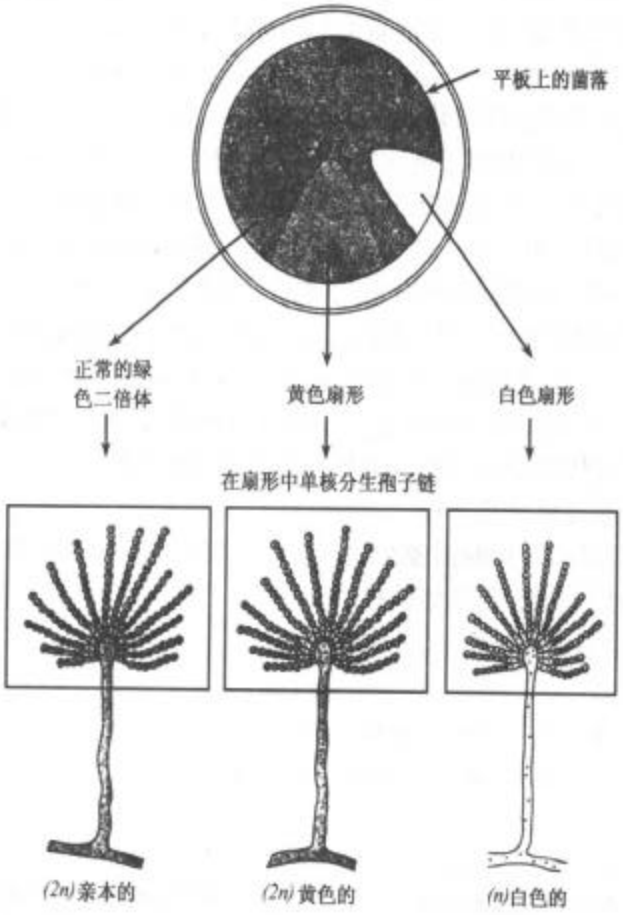
等位基因 w、ad、pro、$pabay$、bi 对其野生型等位基因来说是隐性的 ad、$paba$、bi。表示腺嘌呤(adenine)、对氨基苯甲酸（para-aminobenzoic acid)和生物素(biotin)。这两个品系在基本培养基上都不能生长，但异核体由于基因的互补是可能生长的。

表 5-4　用于曲霉实验的单倍体和二倍体基因型

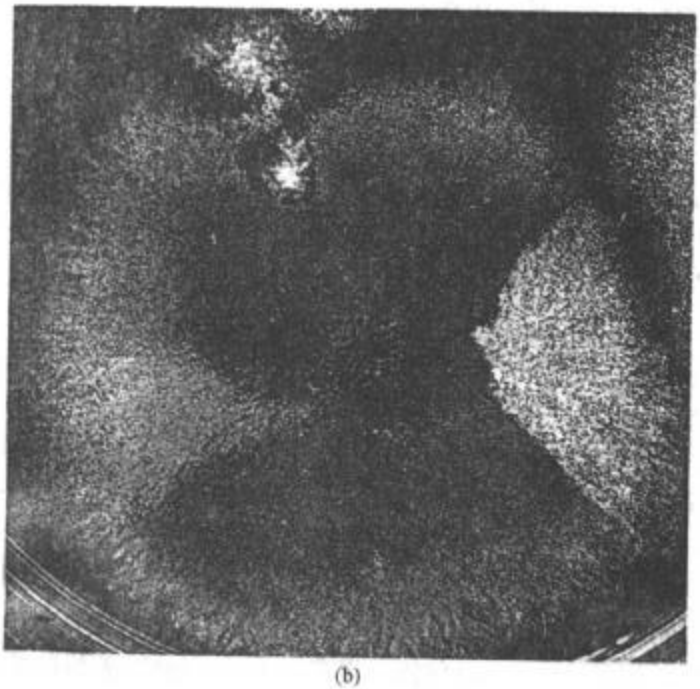
表型		基因型	
单倍体		二倍体	
绿色	w^+/y^+	w^+/w^+	y^+/y^+
		w^+/w	y^+/y^+
		w^+/w^+	y^+/y
黄色	w^+y	w^+/w	y/y
		w^+/w^+	y/y
白色	wy^+	w/y	w^+/w^+
	yy	w/w	y^+/y
		w/w	y/y

PeterJ . Russell : Genetics , 3 rd , 1992 , Table 6.1

w 和 y 等位基因是控制无性孢子的颜色和菌落的颜色，w 导致产生白色的孢子，y 导致产生黄色的孢子。它们都是隐性突变基因。基因型为 w^+y^+ 的品系是绿色的，wy^+ 品系为白色，w^+y 品系为黄色，wy 品系由于上位效应也是白色的。品系 1 和 2 都不是绿色的。然而异核体通常有黄色和白色的孢子混合，而且也有杂色孢子的存在（图 5-17)。它们的颜色是由其基因决定的。而单倍化后的孢子也有绿色，不难看出这一定是重组的结果(表 5-4)。在单倍体中将有一半是白色的，若将白的分离出来，利用选择培养基就可以测出有的位点和亲本的基因型不同，如 pro 位点，原来白色亲本是 pro^-，也就是在培养基中一定加入 Pro 才能生长，而分离出的白色单倍体却有一部分



(a)（黄色的也可能是单倍体，白色的也可能是二倍体）



(b)

图 5 - 17　构巢曲霉有丝分裂分析。（转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 24 - 4）

为 *pro*⁺，即不需加 Pro 就可以生长，这当然是有丝分裂重组的结果，我们将重组类型的菌落数除以总的菌落数就可以得知这两个位点（*w* 和 *pro*）之间的距离。

第六节　人类染色体作图

人类染色体作图也就是基因定位，随着分子生物学的飞速发展，定位的方法也越来越先进。到 1988 年为止人类基因图工作组（HGM）公布在常染色体上已确定 480 个位点，在 X 染色体上已确定了 139 个位点。

定位的方法有体细胞杂交，家系连锁分析，原位杂交，剂量效应法，染色体畸变，限性酶的精确分析等，其中前三者用得较多。

一、体细胞杂交法

人类的染色体和小鼠的染色体无论是数目还是形态都不相同，通过细胞融合可以得到人和小鼠的杂种细胞，这种细胞不稳定、在有丝分裂的过程中总是要排斥人类的染色体，最终仍随机地留下几条人类的染色体，形成不同的细胞系或"克隆"，这才稳定下来，通过 **Giemsa 分带**或其他分析方法，人们能很容易地鉴别出各细胞系中留下的人类染色体是第几号染色体，这就给基因定位提供了前提。我们可通过测定酶的活性来确定那些小鼠不含有而人类却具有的酶的基因座位，以及小鼠为突变型，人类具有其野生型等位基因的酶的位点。例如肽酶 C（peptidase C）是小鼠中没有的，而人类是可以产生此酶。若在杂种细胞系中除小鼠的全套染色体组外还留有一条人类 1 号染色体，同时又测得肽酶 C 的活性，那就可以推测，肽酶 C 的基因位于人类的 1 号染色体上。再如胸苷激酶（TK）是小鼠和人类都具有的酶，那么我们可以选择 *TK*⁻ 小鼠的细胞和人类的细胞融合，用 HAT（含有次黄嘌呤 H，氨基喋呤 A 和胸苷 T）的培养基来选择杂种细胞，只有具有 *TK* 酶活性的杂种细胞才能存活，那么这个酶的基因一定位于杂种细胞中留有的人类染色体上。但杂种细胞留有的人类染色体常常并不止一条，这样就要通过多个相关的杂种细胞系来加以比较分析（如表 5－5），从表中我们不难推断 *TK* 位于人类的第 17 条染色体上。确定一个基因要多个相关克隆并非易事，因为留下的染色体是随机的，所以获得的细胞系不一定是我们所需要的，为此人们采用了**克隆嵌板法**（clone panel method）来筛选就要方便得多，虽构建克隆嵌板并不容易，但一经建立便一劳永逸，事半功倍。

表 5－5　杂种细胞系的染色体和 *TK* 活性

细胞系	人类染色体	*TK*活性
(1)	5,9,12,21	－
(2)	3,4,17,21	＋
(3)	5,6,14,17,22	＋
(4)	3,4,9,18,22	－
(5)	1,2,6,7,20	－
(6)	1,9,17,18,20	＋

表 5－6　克隆嵌板

细胞系	残留的人类染色体							
	1	2	3	4	5	6	7	8
A	＋	＋	＋	＋	－	－	－	－
B	＋	＋	－	－	＋	＋	－	－
C	＋	－	＋	－	＋	－	＋	－

如表 5－6 所示，如果有三个不同的细胞系，每个细胞系含有四条染色体，A 系含有 1、2、3、4 号，B 系

含有 1、2、5、6号，C系含有1、3、5、7号，那么如果三个细胞系都具有某个酶的活性，而A和C都不具有，那么这个酶的相应的基因一定位于1号染色体上；若仅B具有某个酶的活性，那么此酶的基因可能在6号染色体上；依此类推就可以比较快速准确地进行基因定位。现已分离了仅具有一条人类染色体的中国地鼠杂种细胞，毋庸置疑，此将加快人类的基因定位。

二、原位分子杂交

细胞融合的方法只能将某个基因定位于某条染色体上，但却无法确定其具体的位置。用玻片**原位分子杂交**就可以将已知的基因定位于染色体的一定区域。方法是这样的。首先要制备较好的人类中期染色体片，经变性处理使其DNA双链打开，然后将同素标记或带有荧光标记的基因探针（probe）和玻片上的DNA进行分子杂交，经放射自显影或在荧光显微镜下观察摄影，就可以确定该基因在染色体上的具体位置。如人类的胰岛素基因就是用这种方法定位在11号染色体的1区5带上。

三、家系分析法

家系分析法是通过分析家系中连锁基因的重组来确定同一条染色体上基因的排列顺序以及两个基因之间的遗传距离。应当说比原位分子杂交的结果更为精确，但由于人类的家系相对比较小，世代很长，同时不能像动物那样进行杂交实验，因此这种方法的使用受到了很大的局限，一般较适合X连锁基因的定位，直到20世纪60年代后期人们一直都采用这种方法定位。

其他的一些定位方法如细胞学方法和基因剂量效应法是应用于一些特殊的情况。如1968年约翰霍普金斯大学的学生Donahue在做染色体实验时发现自己的1号染色体缢痕区增长，通过对他自己家族染色体的观察，进一步发现这一异常是遗传的，并与特殊血型（Fy）具有平行性，因此将血型基因定位于1号染色体上，这是首次定位常染色体上的基因，从而建立了细胞学定位法。1973年发现第二条染色体短臂缺失（$2P^-$）的患者红细胞酸性磷酸酶（ACP－1）活性明显降低，于是将此酶定位于第2条染色体上，这就是利用剂量效应的方法。

Dulbecco于1986年首次提出了“人类基因组工程”，以期能确定人类及几种模型生物整个基因组的序列，得到了很多著名科学家的支持，这一项浩大的国际协作研究计划终于在1990年正式开始实施。据估计人类的基因组约含30亿个碱基对，有10万个基因分散在基因组中，其中5000个已被编目，1900个已被定位，600个已被克隆分离出来。一旦这项宏伟蓝图得以完成，那么基因定位将会迎刃而解。

第六章

数量遗传

基本概念

非连续性状在生物的表型中只占少数，而连续性状较为广泛。

连续性状有多种表型，此是由于它们由许多基因型所编码或由于环境因子的影响，使每一种表型产生了一个表型范围。

遗传的多基因假说提出：数量性状是由许多基因决定的，这些基因对表型的影响是可以累加的。

广义遗传力是指一群体的遗传方差和表型方差的百分比；狭义遗传力是群体的加性方差和表型方差的百分比。

由于选择在后代中产生的性状变异称为选择效应(response to selection)。选择效应的大小取决于选择的差异和狭义遗传力。

遗传的相关(correlation)是指受到同一个基因影响的两个性状。当一个性状被选择了，通常相关性状也将显示选择效应。

近交(inbreeding)是指亲缘关系较近的个体之间的交配。近交将导致纯合体的增加。

第一节　数量性状的特点

前面所谈到的一些生物的性状基本上都是质量性状，而在自然界中还广泛地存在另一类性状叫做数量性状。如人类的身高、体重、肤色的深浅、智商等；植物的茎杆高矮、籽粒的重量、种皮的颜色、开花期、棉花纤维的长度、奶牛的泌乳量等等。这些性状和质量性状不同，我们从表 6-1 中就可以看出数量性状都是由多对因数基因控制的，而每对基因的作用是微小的，而多对基因的共同作用就决定了表型特征。(在一个群体中数量性状的变量呈正态分布的如人的身长)对质量性状来说虽然与其相关的生化反应链涉及到多对基因，但只要有其中一对基因发生突变就会引起明显的表型变异，所以常常说是由单基因控制的；数量性状的变异及表型的分布是呈一种正态分布，表型是连续的，也就是说在一个的群体中，若将各个体的数量性状(如体重)及占群体中的比率作图，可以呈现一正态分布曲线(图 6-1)，由低向高逐个排列，相邻个体之间的差异很小，变异呈连续分布，不像质量性状，呈不连续的二项分布。如白化症患者和正常人之间界限分明，并不存在各种中间状态。多基因性状受环境的影响较大，就以身长体重来说便和营养状况直接相关，而质量性状有的也受到内外环境的影响，但影响较小或不受影响，如血型，一旦基因型确定之后不论环境如何改变，血型都是不变的。数量性状由于是多对基因作用的结果，每对基因的作用是符合孟德尔法则的，但多对基因作用的累加使得遗传的规律变得十分复杂，不再符合孟德法则，而必须用统计学的方法来加以研究。数量性状的“量”常常是可以用一定的方法来量度，如身长、体重、智商等，而质量性状的“质”常不易度量，只能定性，如血型等。数量性状常用统计学方法来加研究，所以研究的对象是群体，对于个体来说就难以确定；而质量性状既

可以通过统计来观察他们是否符合孟德尔定律，同时也可以通过基因突变的检测来确定一个个体是否属于正常，是杂合体还是突变体。孟德尔在进行豌豆杂交实验时十分注重研究性状，选择那些特征明显而稳定的性状，如花的颜色、子叶的颜色等，舍弃那些不易区分和不稳定的性状，如叶片的形态和数量等，他所选择的性状实际上是质量性状，而舍弃的是数量性状，从而使得他清晰地发现了基本遗传规律，而 Galton 用"相关"与"回归"的方法研究了人类家庭成员的智力，身高，臂长等性状，提出**融合遗传论**，他所选择的性状正是数量性状，得到的结论只适合于数量遗传而不适合于质量性状。数量遗传学对动植物育种和人类群体遗传来说都是十分重要的基础理论。但几十年来没有什么大的进展，出现了瓶颈现象，正期待着新方法新思路的出现。

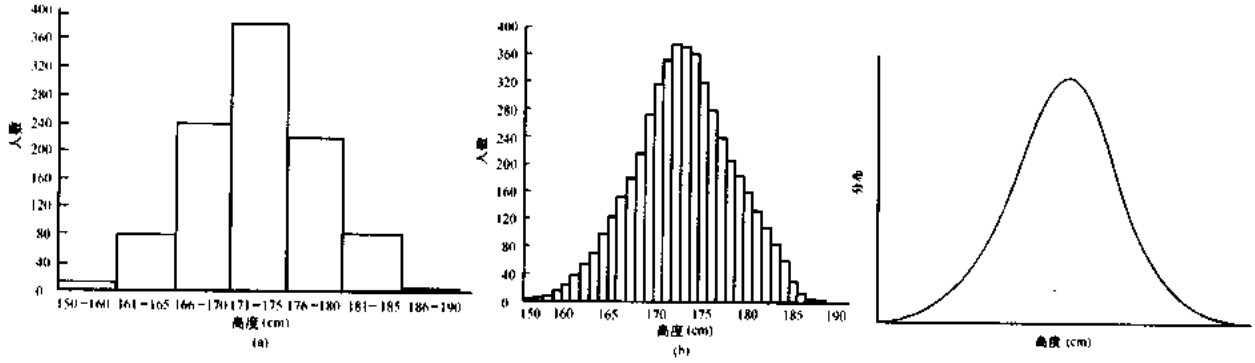


图 6－1　男子身长频率的分布。(a)级差为 5-cm 的柱形图；(b)基差为 1-cm 的柱形图；(c)连续分布。
(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 24－4)

表 6－1　数量性状和质量性状的区别

	基因控制	变异分布	表型分布	受环境影响表	遗传规律	性状特点	研究对象
数量性状	多基因	正态分布	连续	大	非孟德尔遗传	易度量	群体
质量性状	单基因	二项分布	分散	小	孟德尔遗传	不易度量	个体和群体

连续变异和非连续变异

1760 年 Kölrenter 将高株烟草和矮株的烟草进行了杂交，孟德尔也曾将高株豌豆和矮株的豌豆进行了杂交，但二者的结果不同(图 6－2)。Kölrenter 杂交 F_1 代的高度为亲本的平均值(图 6－2 b)；而孟德尔杂交的 F_1 代高度和显性亲本相同(图 7－2 a)。Kölrenter 杂交 F_2 代的平均值和 F_1 代相同，但变异范围增大，呈正态分布。而孟德尔杂交的 F_2 代却产生高度不同的二种后代：高株和矮株，其分离比为 3:1，每一种后代的高度变化范围很小，呈现很窄的峰。这两种杂交虽然都是以高度为研究的性状，但结果显然不同，从而提示这两种植物的高度遗传机制是不同的，Kölreuter 杂交实验中烟草植株的高度可能是由多对基因控制的。

1. 多基因假设

多基因学说可以追溯到 1909 年 Hermann Nilsson-Ehle 的小麦种皮颜色的研究，他

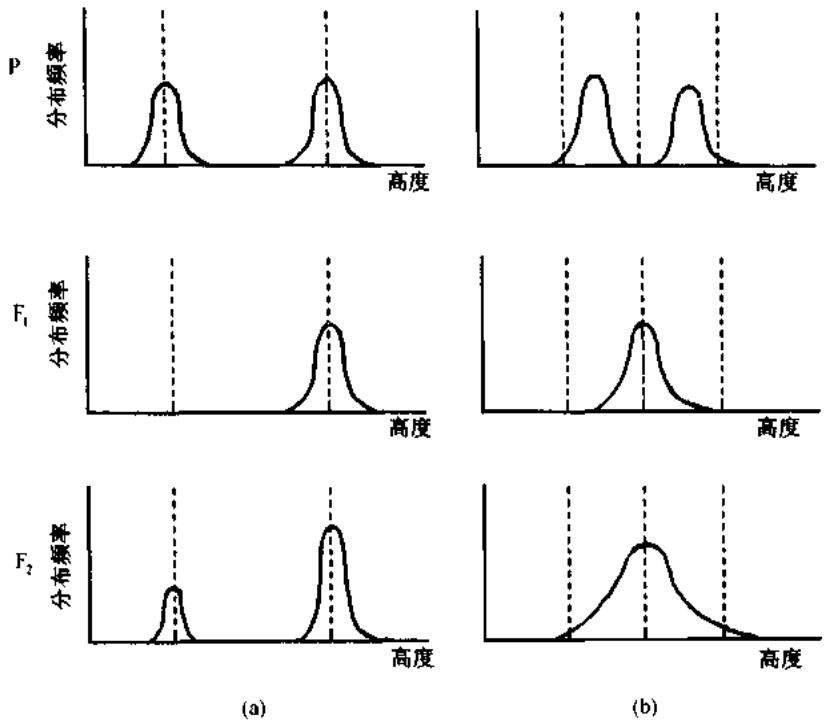


图 6-2 植株的高矮。
有的属于质量性状(a)有的属于数量性(b)二者的遗传规律是不同的。

将红色种皮和白色种皮的纯种小麦进行杂交，获得的 F_1 代小麦表型相同都是红色的，但颜色比红色亲本的要浅，介于两亲本之间。这一现象并不能排除不完全显性，但当他将 F_1 互交后，情况就不同了。F_2 并未出现深红：粉红：白色 = 1：2：1 的孟德尔分离比，而是出现了 15(红)：1(白)的比率，与单基因的不完全显性有很大的偏差。他将 F_2 代的表型细分一下，发现 F_2 代有 4 种不同的红包和一种白色，其比率为深红：大红：中红：浅红：白色 = 1：4：6：4：1，并注意到 F_2 代中各有 1/16 的后代表型分别和两亲本完全相同。

如何来解释这一结果呢？让我们假设有两对不连锁的基因，红色(R, red)和深红(C, crimson)控制产生红色素，而相应的等位基因 r 和 c 不能产生红色素，以上的杂交组合及 F_1、F_2 代的产生是

P　　深红 (*RRCC*) × 白色 (*rrcc*)

↓

F_1　　中红(*RrCc*)

↓(互交)

	深红	大红	中红	大红	中红	浅红	中红	浅红	白
F_2	*RRCC*	*RrCC*	*rrCC*	*RRCc*	*RrCc*	*rrCc*	*RRcc*	*Rrcc*	*rrcc*
	1/16	2/16	1/16	2/16	4/16	2/16	1/16	2/16	1/16

若 R 对 r，C 对 c 是完全显性的话，后代应只有 4 种表型 RC、Rc、rC 和 rc，比率为 9：3：3：1。这与实验所得到的结果不符，所以假设不能成立。

我们再注意一下实验结果的表型比 1：4：6：4：1 和 $(a+b)^4$ 的各项系数相同，

那么一个简单的解释就是红色素合成的深浅是基因剂量控制的，即由 *R* 或 *C* 的数目决定，*RRCC* 相当于二项式展开后的 a^4 是深红的；*rrcc* 则相当于 b^4 是白色的，其他各项不同表型和 *R* 或 *C* 的多少正好一致。*R*、*C* 这种能产生色素的等位基因对表型有贡献就称为**贡献等位基因**（contributing alleles），而 *r*、*c* 不能产生色素就为**非贡献等位基因**（noncontributing alleles），因此小麦种皮颜色的遗传属于 4 个贡献等位基因的多基因遗传。

表 6－2　多基因控制的数量遗传中等位基因数目和基因型、表型数及分离比的关系

等位基因对的数目	分离的等位基因数	F_2 中性状极端表达的比率	F_2 中的基因型数	F_2 中的表型数	F_2 各表型比为二项式各项系数
1	2	$(1/4)^1 = 1/4$	$(3)^1 = 3$	3	$(a+b)^2$
2	4	$(1/4)^2 = 1/16$	$(3)^2 = 9$	5	$(a+b)^4$
3	6	$(1/4)^3 = 1/64$	$(3)^3 = 27$	7	$(a+b)^6$
4	8	$(1/4)^4 = 1/256$	$(3)^4 = 81$	9	$(a+b)^8$
n	2n	$(1/4)^n$	$(3)^n$	2n + 1	$(a+b)^{2n}$

在不同品系的小麦杂交中，有的 F_2 只有三种不同表型：深红，中红和白色，红与之比为 3∶1；也有的 F_2 代红与白之比为 63∶1。而 3∶1 的情况可能是带有 2 个贡献基因的单个基因系统。63∶1 可能是带有 6 个贡献基因的三对等位基因系统，各种表型的比率是 $(a+b)^6$ 的各项系数的比为 1∶6∶15∶20∶15∶6∶1。

根据小麦种皮颜色的遗传初步提出了多基因学说：

(1)决定数量性状的基因是一系列的贡献等位基因和非贡献等位基因；

(2)各贡献等位基因对某种性状的作用是相同的；

(3)各贡献等位基因的作用是可以积加的；

(4)在多基因遗传的各位点之间无上位效应；

(5)在多基因遗传中的各基因是不连锁的；

(6)没有环境的影响。

以上假说的后三点并不是真实的，在多基因遗传中可能有上位效应和环境的影响，也不能排除连锁的可能，因此这个学说还必须得到修正。

2. 玉米穗长实验

玉米是遗传学和细胞学常用的实验材料，穗长是其性状之一。1913 年 Rollins Emerson 和 E. East 报导了他们对玉米穗长的研究结果，他们将甜玉米（平均穗长 6.63cm）和爆玉米（平均穗长为 16.80cm）进行杂交，这两种玉米各自的穗长变异都不大，平均长度各为 6.63cm 和 16.80cm，杂交结果 F_1 代的平均长度为 12.12cm，处于两亲本长度的平均值之间。亲本是纯种，我们可以假设控制的长度的基因是纯合的，所有两亲本的穗长差异一定是遗传的差异。而 F_1 是杂合体，但基因型相同，那么穗长的差异一定由于环境的差异所引起。

F_2 代的平均长度是 12.89 和 F_1 代大致相似，但 F_2 代的变异范围比 F_1 要大得多（表 6－3），它们的标准差 s 就可在显示这个问题。长穗的 s = 0.816；短穗的 s = 1.887，F_1 的 s = 1.519 而 F_2 的 s = 2.25 2，这些数据表明 F_2 的变异范围最大。

表 6-3 玉米穗长的遗传

穗长(cm) / 频数(个) / 世代	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	合计	平均数 ($\bar{X}$)	方差 (v)	标准差 (s)
P_1	4	21	24	8														57	6.632	0.666	0.816
P_2									3	11	12	15	26	15	10	7	2	101	16.802	3.560	1.887
F_1					1	12	22	47	19	4								69	12.116	2.307	1.519
F_2	1	10	9	6	47	73	68	89	63	25	15	9	1					401	12.888	5.074	2.252

环境因素引起的变异是什么呢?如果说代和 F_1 代的变异反映了环境的影响,那么这种影响同样也作用于 F_2 上,但如果种在同一环境中,我们就有理由假设环境对 F_2 的作用应和 P 代及 F_1 相似。可是数据表明 F_2 穗长变异范围明显增加,这是什么原因引起的呢?比较合理的解释是 F_2 中存在着更大的遗传变异。

玉米穗长的研究资料除了环境的影响以外说明了以下 4 点:① F_1 数量性状的平均值介与两亲本的平均值之间;②F_2 数量性状的平均值和 F_1 相等;③F_2 的变异范围大于 F_1;④ F_2 数量性状的极值比 F_1 更为深入地延伸到两亲本的分布范围中。这显然和孟德尔的单基因性状的遗传是不同的。如何来解释数量性状的遗传规律呢?East 根据玉米穗长和烟草花冠长度遗传的研究发展了多基因学说,提出:

(1) 数量性状受多基因支配;

(2) 这些基因对表型影响小,相互独立,但以积累的方式影响相同的表型;

(3) 每对基因常表现为不完全显性,分离比符合孟德尔法则;

(4) F_1 介于两亲本的中间,F_2 有较大的变异范围。

3. 多基因的分离与连锁

多基因假说有两个基本特点:①每对基因按孟德尔法则遗传;②它们对性状的作用相似,效应较小。因此在表型分布中不能分解出非连续性,于是在环境等非遗传因素的作用下,非连续的,定量的基因型变异就形成了连续的表型分布。

既然这些基因不能用同简单的孟德尔方法来分析,那么怎样才能证实他们是按孟德尔方式遗传?怎样明确这种多基因是否存在连锁呢?

间接的证据是数量性状遗传正反交的结果相同,即使有时稍有差异这在质量性状的遗传中也可见到。正反交结果相同,表明双亲对子代基因型的贡献相同,而且都来自核基因。用另一些方法可以检测出核基因的分离和连锁。表 5-3 中,明显可以看出在数量性状遗传中 F_2 的基因型数和质量性状遗传中 F_2 的基因型数相同,表明分离规律是相同的。多基因学说是建立在多基因不连锁的前提之上,虽然是多基因完全有可能位于同一条染色体上,1923 年 Sax 用菜豆(*Phasolus unlgaris*)杂交证实了多基因和

主基因的连锁。他用一种大粒紫色品系和小粒白色品系杂交，结果显示豆子的颜色符合孟德尔遗传，F_1 代为紫色，F_2 代紫色与白色豆子之比为 3∶1，可见紫色（P）对白色（p）是完全显性，但豆子的大小却呈连续变异，其遗传规律符合多基因理论。可是豆子的色素的基因型又和其大小显示了明显的相关（表 6-4）由于别的实验排除了粒重是 P 基因的多效应这一可能，从而说明在菜豆中控制色素形成的主基因 P 和控制粒重的多基因部分连锁。

表 6-4　菜豆 F_2 代的颜色和粒重

株数	基因型	表型	平均粒重(cg)
45	PP	紫色	30.7±0.6
80	Pp	紫色	28.3±0.3
41	pp	白色	26.4±0.5

第二节　分析数量性状的基本统计方法

一、均值

一个表型分布的状况可以通过二个统计量来表现，那就是均值（Mean）和方差（Variance）。均值又称为平均值（average）。在样本中均值给我们提供了有关表型分布中心的信息。均值可以用个体量度值的总和除以个体数 n 来求得。

$$\bar{X}=\frac{x_1+x_2+x_3+\ldots+x_n}{n}$$

也可以写成

$$\bar{X}=\frac{\sum x_i}{n}$$

X_i 表示样本中各个体的量度值。如果在样本中有的个体量度值相等，则可以合并为一组，其量度值以 m_i 表示，组内的个体数或占总体样本的频数以 f_i 表示，根据表 6-5 的数据则均值可表示为：

$$\bar{X}=\frac{\sum f_i m_i}{n}$$

$$T(\text{two})=x+y \qquad n=20$$

$$\bar{X}=\frac{\sum x_i}{n}=6.65$$

$$\bar{Y}=\frac{\sum y_i}{n}=7.05$$

$$\bar{T}=\frac{\sum f_i m_i}{n}=0.05(11)+0.15(12)+0.20(13)+0.35(14)+0.15(15)+0.10(16)$$

$$=13.7$$

二、方差

方差是表型分布的关键信息，它能反应分布范围也就是个体和均值的离散程度，均值相同的两个样本，它们的方差不一定相同。玉米穗长 F_1 和 F_2 穗长的均值相近，但变异范围明显不同，方差就可反映这一特点，方差是变数$\bar{X}$和平均数的偏差平均平方和。方差公式是：$S^2 = \dfrac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{n-1} = \dfrac{\sum(x_i-\bar{x})^2}{n-1}$

以上是样本方差，若是总体方差分母为 N。每个个体和均值的离差 $(x_i-\bar{x})$ 都有了，为什么还要进行平方变成方差呢？因为我们要求的是各项离差的和，而各项离差有的是正值，有的是负值，相加后可能会互相抵消而产生误差，为了使各项都就成正值，所以进行平方。

因为 $\bar{x}$ 和$(x_i-\bar{x})$不是整数不便计算，所以将上式进一步改为下式：

$$S^2 = \frac{\sum(x_i-\bar{x})^2}{n} = \frac{\sum(x_i^2-2x_i\bar{x}+\bar{x}^2)}{n} = \frac{\sum x_i^2-\sum 2x_i\bar{x}+\sum\bar{x}^2}{n-1} \tag{1}$$

因为 $\bar{x} = \dfrac{\sum x_i}{n}$　　所以 $n\bar{x} = \sum x_i$ (2)

将(2)代入(1)

$$S^2 = \frac{\sum x_i^2-\sum 2x_i\bar{x}+\sum\bar{x}^2}{n-1} = \frac{\sum x_i^2-\sum\bar{x}^2}{n-1} \tag{3}$$

将 $\bar{x} = \dfrac{\sum x_i}{n}$ 代入(3)

$$S^2 = \frac{\sum x_i^2-n(\frac{\sum x_i}{n})^2}{n-1} = \frac{\sum x_i^2-(\frac{\sum x_i}{n})^2}{n-1}$$

以表 6-5 为例求 T 的方差 S_T^2可以用几种方法求得

$$S_T^2 = \frac{(11-13.7)^2+(12-13.7)^2+(13-13.7)^2+\ldots+(15-13.7)^2+(16-13.7)^2}{20}$$

$$= 1.71$$

或

$$S_T^2 = 0.05(11-13.7)^2+0.15(12-13.7)^2+0.20(13-13.7)^2+0.35(14-13.7)^2$$
$$+0.15(15-13.7)^2+0.10(16-13.7)^2 = 1.71$$

或

$$S_T^2 = \frac{1}{n}\sum T_i^2-\bar{T}^2 = 3788/20-187.69 = 1.71$$

三、标准差

方差虽能反映样本的变异范围但它是经过平方后得出的值，因此其单位和个体量的单位意义已不同，为了使得变异范围的单位和个体量度范围相同，人们就将方差开

方，方根以 S 表示。 $S=\sqrt{\frac{\sum(x_i+\bar{x})^2}{n}}$ ，S 就称为标准差(standard deviation，SD)或叫做准误(standard error)。

表 6-5　雄果蝇左(y)右(x)腿上性梳的齿数

x	y	T	mi	$Fi=mi/n$
6	5	11	1	0.05
6	6	12		
5	7	12	3	0.15
6	6	12		
7	6	13		
5	8	13		
6	7	13	4	0.20
7	6	13		
8	6	14		
6	8	14		
7	7	14		
7	7	14	7	0.35
7	7	14		
6	8	14		
8	6	14		
8	7	15	3	0.15
7	8	15		
6	9	15		
8	8	16	2	0.10
7	9	16		

表 6-5 中的 $S_x=0.9096$，$S_y=1.0722$，$S_t=1.308$ 从这个数据我们可以知道对于总体来说 95% 的果蝇的右边性梳的齿数变异范围不超过 ±0.9096，左边性梳的齿数变异范围不超过 ±1.0722 右边性梳范围变异较小。双腿的性梳总数为异范围不超过 ±1.308 个。从(图 6-3)中可以看出均值和标准差决定了表型分布的特点。

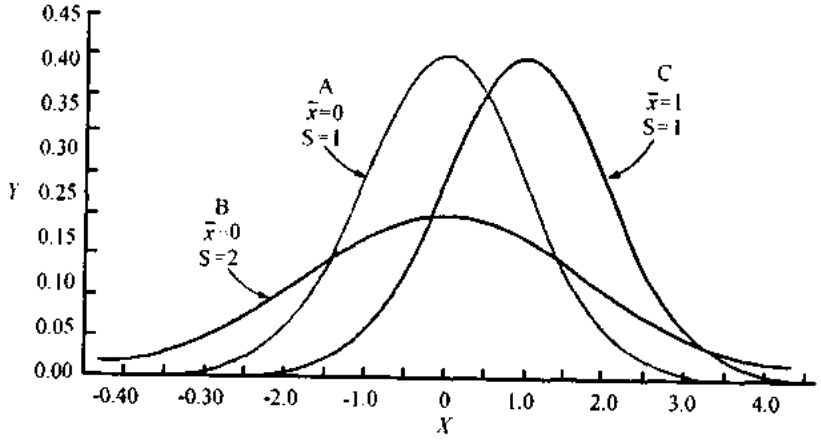


图 6-3　有三个分配函数，其中两个分配函数的均值相同，两个分配函数的标准差相同。
(引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 24-4)

四、相关

在我们研究的各种性状中，常常有二个变量是有联系的或相关的，也就是说一个变量发生变化，另一个变量也发生变化。比如大部分动物头的大小和身体的大小是相关的。两个或几个统计变量同时变化的密切程度用相关系数（correlation coefficient）来表示，其域值范围为 $-1 \leqslant \gamma \leqslant 1$，它同时也指出相关的方向。最小的相关系数 -1 表示绝对负相关；最大的相关系数 +1 表示正相关。假设我们有两个统计变量 x（如身长）和 y（如头的宽度）我们希望能计算 x 和 y 之间的相关（correlation）我们先要获得 x 和 y 的协方差（covariance）。协方差是量度两变数之间的互变关系，γ 在 x 和 y 方向中的离差的积数之和除此之除以自由度就称为 x 和 y 的协方差 COVxy：

$$\mathrm{COV}xy = \frac{\sum(x_i + \bar{x})(y_i + \bar{y})}{n-1}$$

为了便于计算，上式可写成

$$\mathrm{COV}xy = \frac{\sum x_i y_i - \frac{1}{2}(\sum x_i y_i)}{n-1}$$

相关系数 γ 等于 xy 的协方差除以二者标准差的乘积

$$\gamma = \frac{\mathrm{COV}\ xy}{SxSy}$$

现以虎螈的体长和头宽这两个统计量为例（表 5-6），通过计算，$\gamma = 0.91$。

表 6-6 虎螈体长和头宽的相关性

体长(mm)x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	头宽(mm)y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	x_i y_i
72.00	-7.92	62.67	17.00	-0.75	14.06	1224
62.00	-17.92	321.01	14.00	-3.75	14.06	868
86.00	6.08	37.01	20.00	2.25	5.06	1720
76.00	-3.92	15.34	14.00	-3.75	14.06	1064
64.00	-15.92	253.34	15.00	-2.75	7.56	960
82.00	2.08	4.34	20.00	2.25	5.06	1640
71.00	-8.92	79.51	15.00	-2.57	7.56	1065
96.00	16.08	258.67	21.00	3.25	10.56	2016
87.00	7.08	50.17	19.00	1.25	1.56	1653
103.00	23.08	532.84	23.00	5.25	27.56	23.69
86.00	6.08	37.01	18.00	0.25	0.06	1548
74.00	-5.92	35.01	17.00	0.75	0.56	1258
$\sum x_i$		$\sum(x_i - \bar{x})^2$	$\sum y_i$		$\sum(y_i - \bar{y})^2$	$\sum x_i y_i$
= 959.00		= 1686.92	= 213.00		= 94.25	= 17385

$$\bar{x} = \sum x_i / n = 959/12 = 79.92$$

$$\bar{y} = \sum y_i / n = 213/12 = 17.75$$

$$S_x^2 = \sum (x_i - \bar{x})^2/(n-1) = 1686.92/11 = 153.35$$

$$S_x = \sqrt{S_x^2} = \sqrt{153.35} = 12.38$$

$$S_y^2 = \sum (y_i - \bar{y})^2/(n-1) = 94.25/11 = 8.75$$

$$S_y = \sqrt{S_y^2} = 2.93$$

协方差　$$COVxy = \frac{\sum x_i y_i - \frac{1}{2}(\sum x_i y_i)}{n-1} = \frac{17385 - \frac{959 \times 213}{12}}{12-1} = 32.97$$

相关系数　$\gamma = COVxy / SxSy = 32.97/(12.38 \times 2.93) = 0.91$

若相关系数为为正值时，表明当一个变量增加时另一变量也随这增加（图 6-4），在向日葵中若种子的大小正相关于种子的多少。那么产生大种子的植株也会结较多的种子。图中的 b、c、d、f 都是正相关。相关系数为负数时表明当一个变量增加时，另一个变量也随之减少，如果种子的大小和数量是负相关的话，那么产生大种子的植物结的

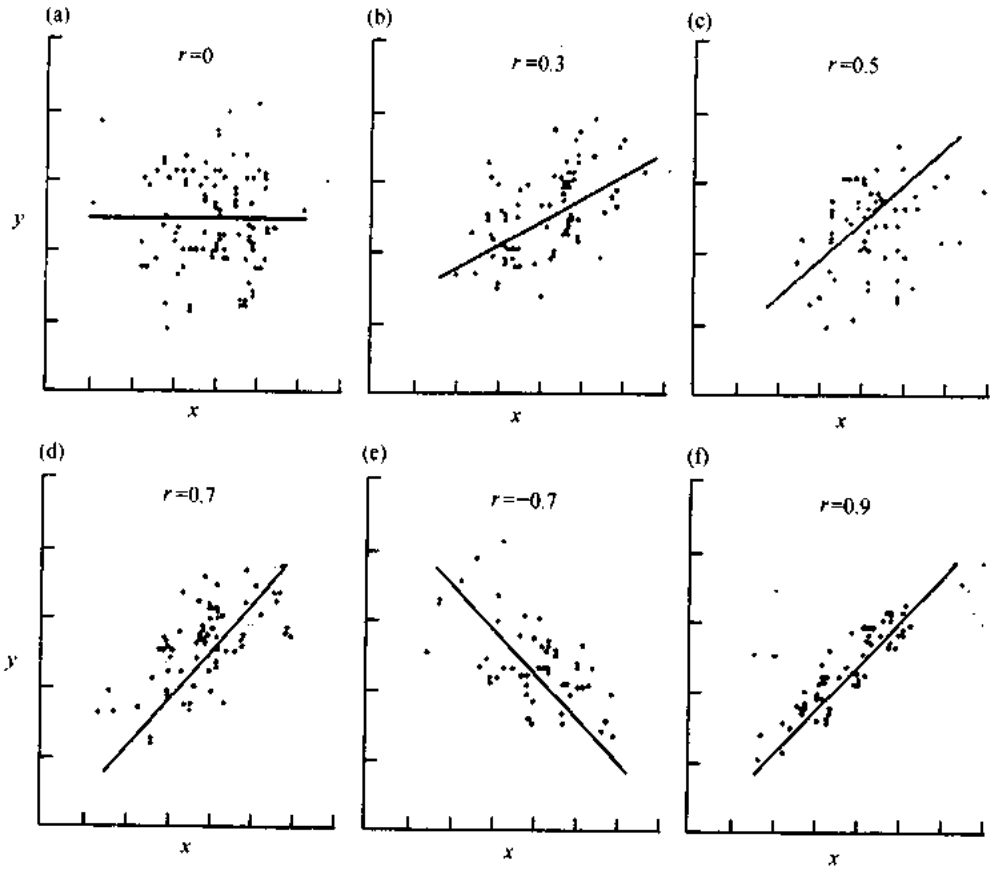


图 6-4　若相关系数(γ)为正值时，表明当一个变量增加时另一变量也随这增加。图中 X 和 Y 变量相关。图 b、c、d 和 f 表示正相关，图 e 表示负相关。γ 为相关系数。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 22-9）

种子要比结小种子的植物来得少。图中的 e 就表示负相关。当相关性系数靠近 -1或 +1时，相关性是强的。即一个变量的改变总随着另一变量的改变。如图 6-4 的 f，$\gamma = 0.9$ 表明相关性很强，相反当 $\gamma = 0$ 时(图 6-4a)表明两个变量之间没有什么相关性。

我们必须注意：①变量之间的相关性仅表示这两个变量有联系，但并不意味着有因果关系的存在；②相关性并不是表明两事件相同。仅意味着一个变量的改变和另一变量中的改变有关联。两个变量可能高度相关，但数值有很大的差别。例如当代大学时代的男子身高是和他们父亲的身高相关的，高个子的父亲产生了高个子的儿子，矮个子父亲生育了矮个子的儿子，这种相关性是由于影响人类身长的基因作用的结果。但现代的大部分男大学生都要高于他们的父亲，这可能是由于近年来较好的营养和健康状况增加了所有个体的身长。这样父亲和儿子在身长上显示了相关性，但他们的身长并不相同。

五、回归

相关系数告诉我们变量之间的联系是强是弱，并表明这种关系是正是负。但关于变量间的精确的关系我们凭相关系数确难以确定。我们常常想知道一个变量有多少变化是和另一个变量的变化有关联。比如前面我们所举的父子身高的例子中我们要问如果父亲是 170cm 高，儿子可能有多高呢？回答这个问题是要用另一种统计方法，回归(regression)分析。

"回归"这个词是由 F.Galton 创立的。回归是指因变量对自变量的依存关系。相关的两个变量之间是没有从属关系的，而回归关系的两个变量是有从属关系的。比如父子的身长存在着回归的关系，高个子的父亲常生高个子的儿子，但儿子身长超过父亲的概率要小于低于父亲的概率；同样矮个子的父亲常生矮个子的儿子，但儿子身长比父亲更矮的概率要小于比父亲高的概率。也就是说不可能无限制的一代比一代高，或一代比一代矮。后代的高度有向平均高度回归的趋势；离开均值越远，所受到回归的

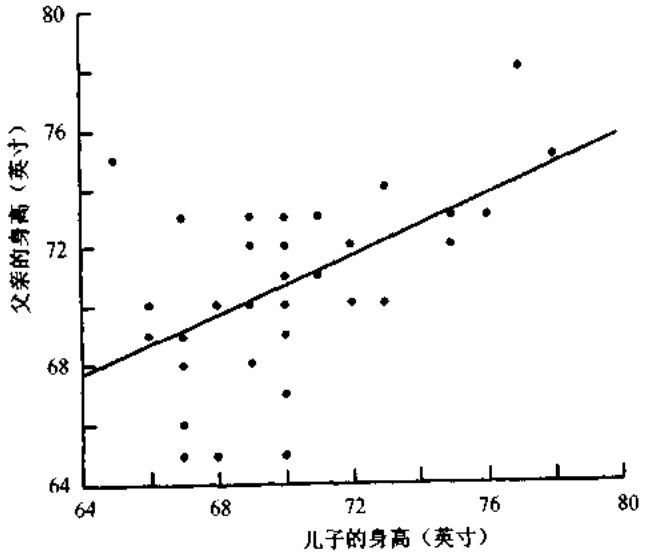


图 6-5　儿子身高对父亲身高的回归回归方程是 $y = 36.05 + 0.49x$。
(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 22-9)

压力也越大。“回归”这个词就由此而来。

两个可变量之间的关系可以用回归线(regression line)来表示。图 6－5 表示父亲和儿子身高的关系。横坐标表示父亲的高度，纵坐标表示儿子高度，每一点表示标本中父亲和儿子高度的交点。回归线是数学估计线，表示最适合的点。

回归的一般意义是表示一个因变量能随着另一个自变量的改变而变；通过回归计算可根据自变量的值以一定的置信程度估计相应的因变量的值。另一个重要的意义是若引入自变量，可以使因变量的偶然性内部变差缩小的话，那么我们说在这两种变量间存在着回归的关系。例如以成年人的身高作为因变量，单独测量这个变量就会发现其本身变差很大，若将年龄作为自变量引入，再把身高按年龄的大小依次分组时，发现年龄组内的身高变差比不分组前的变差要小得多。因此由年龄的引入成功缩小了身高的变差，我就可以确定身高在年龄上存在回归。回归线是一条直线，所以我们可以用直线方程来表示。

$$y = a + bx$$

a 是 y 的截距，b 是斜率，又称为回归系数。回归系数如何计算呢？我们首先考虑下列法则：把双变量的随机样本在直角坐标上作成散点图，必定存在这样一根直线，到每个点的 y 方向的距离的平方和为最小，也就是“最小二乘”法则。只要散点存在，那么这根直线就可以作出，且只有惟一的一条。这就是我们所求的回归线。回归系数 b 可以按下式求得

$$b = \frac{COV_{xy}}{S_x^2}$$

约去自由度　$b = \dfrac{\sum(x_i + \bar{x})(y_i + \bar{y})}{\sum(x_i - \bar{x})^2}$ 为便于计算，可改写为

$$b = \frac{\sum x_i y_i - \dfrac{(\sum x_i)(\sum y_i)}{n}}{\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}}$$

第三节　遗传力

数量性状受到环境因素的影响很大，那么表型的变异可能有遗传的因素，也有环境的因素，甚至还有环境和遗传相互作用的因素。例如 15℃时基因型 *AA* 的植物平均高度为 40cm，而基因型 *Aa* 型的植物仅有 35cm 高；但是在 30℃时情况就不同了，*AA* 植株的平均高度为 55cm，而 *Aa* 型植株反长得更高，为 60cm。同一种基因型在不同的温度下表型不同，这一变异是由环境引起的，所产生的方差称环境方差(V_E)；在同样的温度下，不同基因型的高度不同，这一表型变异是遗传因子所引起的，所产生的方差称遗传方差。在任何的温度下，植物的表型即受到温度的影响，又受到遗传因素的影响，甚至遗传因素和环境相互作用也会影响到植物的表型，这种影响产生的方差称环境和遗传的协方差。

表型是由遗传和环境共同作用的结果，可以下式表示

$$P = G + E$$

若以方差表示 $S_p^2 = S_g^2 + S_e^2 + 2COV_{ge}$

对于一个表型的变异究竟是遗传因子起主要作用还是环境因子起主要作用呢？为了解答这个问题人们又引入一个概念叫**遗传力**（heritability）。遗传力表明某一性状受到遗传控制的程度。它介于0与+1之间，当等于1时表明表型变异完全是由遗传的因素决定的，当等于0时表型变异由环境所造成，从表5-7中我们不难看出，人类的身长，智商及精神分裂症中的一种类型受遗传控制的程度较大；相反像数学的天赋等遗传的作用很小，主要是依靠后天的努力和培养。遗传力又分为广义遗传力（broad-sense heritability）和**狭义遗传力**（narrow-sense heritability）。广义遗传力是指表型方差（V_P）中遗传方差（V_G）所占的比率：

$$H^2 = \frac{V_G}{V_P}$$

表6-7 人类一些性状的遗传力

性状	遗传力	性状	遗传力
身材	0.81	理科天赋	0.34
坐高	0.76	数学天赋	0.12
体重	0.78	文史天赋	0.45
口才	0.68	拼写能力	0.53
IQ(Binet)	0.68	先天性幽门狭窄	0.75
IQ(Otis)	0.80	精神分裂症	0.80
唇裂	0.76	糖尿病	0.75
糖尿病	0.75	高血压	0.62
高血压	0.62	冠状动脉病	0.65

狭义遗传力是指表型方差（V_P）中加性方差（V_A）所占的比率：

$$H^2 = \frac{V_A}{V_P}$$

什么是加性方差呢？

在数量性状中，每对等位基的作用有积加的效应，如控制高度的基因，一个等位基因 *g* 平均可贡献2cm的高度，*G* 平均贡献4cm的高度，*gg* 纯合贡献4cm高度，*Gg* 杂合可贡献6cm高度，或者7cm，8cm的高度，*GG* 纯合可贡献8 cm的高度，它们的遗传方差称为**加性遗传方差**（additive genetic variance）。若 *G* 对 *g* 有显性作用，那么 *Gg* 将贡献8 cm，显性作用使表型产生的方差称显性方差。若有非等位基因之间存在上位效应的话还必须考虑到上位方差，或称为**互作方差**（interaction variance）V_I，因此遗传方差是由下式表示：

$$V_G = V_A + V_D + V_I$$

总表型方差可以表示为：

$$V_P = V_A + V_D + V_I + V_E + V_{GE}$$

一、遗传力的计算

遗传力是可用不同的方法来计算的。现举例说明。

假设影响长度的一对等位基因 A 和 a，在一个群体中

基因型	aa	Aa	AA
表现型	10	18	20
频 度	0.36	0.48	0.16

a 基因的平均效应　$\bar{a} = \dfrac{2(0.36)(10) + 1(0.48)(18)}{2(0.36) + 1(0.48)} = 13.2\text{cm}$

A 基因的平均效应　$\bar{A} = \dfrac{2(0.16)(20) + 1(0.48)(18)}{2(0.16) + 1(0.48)} = 18.80\ \text{cm}$

$\bar{A}$ 和 $\bar{a}$ 之间的平均差异　$\bar{A} - \bar{a} = 18.80 - 113.20 = 5.60\text{cm}$

当 A 和 a 杂合时 $Aa = 18$，并不等于 AA 和 aa 的平均数，表明有某些显性效应存在。

Gg 杂合状态下的高度 6cm 正好等于两个纯合基因型 GG 和 gg 的平均值，$6 = (4 + +8)/2$ 说明没有显性；当 Gg 贡献 8 cm 高度，和 GG 的表型相同，则表现出了 G 对 g 完全显性。

若 $Gg = (GG + gg)/2 = (4 + 8)/2 = 6$ (cm) 说明 G 对 g 无显性；

若 $Gg = GG = 8$ 表明 G 对 g 完全显性；

若 Gg 的表型为 $6 < Gg < 8$ 则为不完全显性；

若 $4 < Gg < 6$ 则 g 对 G 为显性；

若 $Gg > AA(8)$ 表明超显性。

二、加性方差和显性方差

只有在基因纯合时才能显示相加效应，在杂合时显示显性效应。我们把群体总的遗传方差分成加性方差(additive genetic variance)S_a^2 和显性方差(dominance variance)S_d^2，加性方差是由群体中两种纯合体(AA 和 aa)之间的平均差异而形成的；显性方差 S_d^2 是由基因杂合作用而产生的。这样

$$V_G = V_A + V_D$$

$$S_g^2 = S_a^2 + S_d^2$$

现以 $aa = 10$ $Aa = 18$ $AA = 20$ 为例

$\bar{X} = \sum f_i x_i = 0.36 \times 10 + 0.48 \times 18 + 0.16 \times 20 = 15.44\,(\text{cm})$

总的遗传方差

$S_g^2 = \sum f_i (x_i - x)^2 = 0.36 \times (10 - 15.44)^2 + 0.48(18 - 15.44)^2 \times 0.16(20 - 15.44)^2$

$= 17.13\ (\text{cm}) = \sqrt{d}$

a 等位基因的频率　$f_A = \sqrt{0.36} = 0.60$

或者　$f_a = \dfrac{2(aa) + 1(Aa)}{2} = \dfrac{2 \times 0.36 + 1 \times 0.48}{2} = 0.60$

A 等位基因的频率　$f_A = \sqrt{AA} = \sqrt{0.16} = 0.40$

等位基因的平均方差　$S^2 = f_a(\bar{a} - \bar{x})^2 - f_A(\bar{A} - \bar{x})^2$

$$= 0.60 \times (13.20 - 15.44)^2 + 0.40 \times (18.80 - 15.44)^2$$
$$= 7.525(cm^2)$$

以上是以一个等位基因计算的，而二倍体生物都带有两等位基因，因此，

$$S_a^2 = 2 \times 7.525 = 15.05\ (cm^2)$$

由于 $S_F^2 = S_a^2 + S_d^2$　　所以 $S_d^2 = S_F^2 - S_a^2 = 17.13 - 15.05 = 2.08\ (cm^2)$

三、总的表型方差

总的表型方差　　$S_p^2 = S_g^2 + S_e^2 = S_a^2 + S_d^2 + S_e^2$

表型方差只有在 F2 代中通过对不同表型的后代的测量方可得到。当我们求得 $Sg2$ 和环境方差 Se^2 时也可求得 Sp^2。

求环境方差的方法一般是测算基因型相同的个体组成的某群体的方差，即是环境方差。因在基因型相同时表型的变异只可能是环境引起的。常常采用的方法是

$$V_E = \frac{V_{P1} + V_{P2}}{2} = S_e^2 = \frac{S_{P1}^2 + S_{P2}^2}{2}$$

Vp1 和 Vp2 表示植物杂交中两个亲本的方差。由于亲本都要求是纯系，基因型可视为一致的。那么每一个亲本产生的表型变异应是环境造成的。为了较为精确，希望有不同的样本，所以采用两个亲本。但由于有些植物纯系表现出明显的退化，如玉米纯化后植株矮小，结实少，这样会带来误差。为了消除或减少误差可用下式来计算：

$$V_E = (V_{P1} + V_{P2} + V_{F1})/3$$

与前式不同之处仅增加了 F_1 的表型方差。因 F_1 虽是杂合的，但基因型相同，如果表型发生变异当然可以推测为环境因素所造成的，同时 F_1 又有杂种优势，比亲本长得高大，结籽率也高，从而解决了性状退化所带来的误差。

我们求得环境方差后很容易就可以求出广义遗传力

$$H_2 = \frac{V_G}{V_P} = \frac{V_G}{V_G + V_E}$$

四、Warner 法

1952 年 Warner 提出用两个回交世代估算狭义遗传力的方法。

设一对基因杂交 $Aa \times aa \rightarrow \frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa$

AA 的表型值为 a，Aa 的观察值为 d，aa 的观察值为 $-a$。

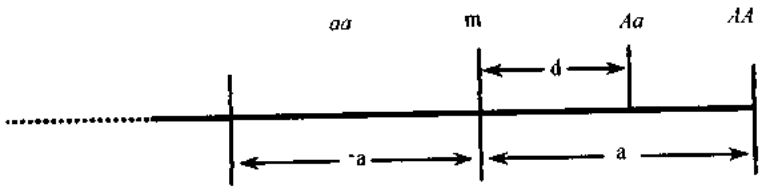


图 6-6　表示 AA、Aa、aa 不同表型计量的模式图。$AA - aa = 2am$ 为 Aa 和 aa 的平均值，杂合体 Aa 和均值之间的数量差异为 d。d/a 为显性程度。

AA 和 *aa* 的差异为 2a，in 点为 *AA* 和 *aa* 的均值，又称中视点。*Aa* 到 m 点的距离为 d，d/a 称优势比率(potence rate)(图 6－6)。

当 d = a　d/a = 1　Aa = AA　时完全显性

d = m　Aa = m　无显性

m < d < a　0 < d/a < 1　部分显性或不完全显性。

d > a　d/a > 1　超显性(superdominance)

m > d > －a　*aa* 对 *AA* 为显性

以上是以一对基因时的情况，当群体性状由多基因控制时，d/a 就不能直接度量显性程度，因为可能有一些 d 是正值，而另一些是负值可以互相抵消。表 6－8 是 *Aa* 自交的结果，但此均值是差异部分的均值，并非是整体的均值。

表 6－8　F_2 的均值和方差计算

基因型	f_i	x_i	$f_i x_i$	$f_i x_i^2$
AA	1/4	a	1/4　a	1/4　a^2
Aa	1/2	d	1/2　a	1/4　d^2
Aa	1/4	－a	－1/4　a	1/4　a^2
合计	1	$\sum x_i = d$	$\sum f_i x_i = \frac{1}{2}d$	$\sum f_i x_i^2 = \frac{1}{2}a^2 + \frac{1}{2}d^2$

表 6－9　B1 的平均数和遗传方差的计算

基因型	f_i	x_i	$f_i x_i$	$f_i x_i^2$
AA	1/2	a	1/2　a	1/2　a^2
Aa	1/2	d	1/2　d	1/2　d^2
合计	1	$\sum x_i = a + d$	$\sum f_i x_i = \frac{1}{2}(a+d)$	$\sum f_i x_i^2 = \frac{1}{2}(a^2 + d^2)$

表 6－10　B2 的均值和遗传方差的计算

基因型	f_i	x_i	$f_i x_i$	$f_i x_i^2$
Aa	1/2	d	1/2　d	1/2　d^2
aa	1/2	－a	－1/2　a	1/2　a^2
合计	1	$\sum xi = d - a$	$\sum f_i x_i = \frac{1}{2}(d-a)$	$\sum f_i x_i^2 = \frac{1}{2}(a^2 + d^2)$

将 *Aa* 与 *AA* 回交，回交 1 代 B_1 的频度和观察值如表 6－9 所示，根据表 6－9，B_1 差异部分的均值

$$\bar{B}_1 = \sum f_i x_i = \frac{1}{2}(a+d)$$

B_1 的方差　$S_1^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = \frac{1}{2}(a^2+d^2) - \frac{1}{4}(a+d)^2 = \frac{1}{4}(a+d)^2$

将 F_1 *Aa* 与 *aa* 回交，回交一代 B_2 的频度和观察值如表所示。B_2 差异部分的均值

$$\bar{B}_2 = \sum f_i x_i = \frac{1}{2}(a-d)$$

B_1 的方差 $S_2^2\sum x_i^2 - \frac{(\sum x_i)^2}{n} = \frac{1}{2}(a^2+d^2) - \frac{1}{4}(d-a)^2 = \frac{1}{4}(d+a)^2$

回交一代的平均方差为 $\frac{1}{2}(S_1^2+S_2^2) = \frac{1}{4}a^2 + \frac{1}{4}d^2$

因 $F_2 = 1/2a^2 + 1/4d^2$ 这儿未考虑环境方差，实际上是遗传方差，其 $1/2a^2$ 是加性方差，即 $Sa^2 = 1/2a^2$；$1/4d^2$ 是显性方差，即 $S_d^2 = \frac{1}{4}d^2$

若将 V_{F2} 的表型方差和回交 1 代平均方差相减

$$V_{F2} - \frac{1}{2}(V_{B1}+V_{B2}) = (\frac{1}{2}a^2 + \frac{1}{4}d^2) - (\frac{1}{4}a^2 + \frac{1}{4}d^2) = \frac{1}{4}a^2$$

即使考虑环境方差的话，双方也可以减掉，并不影响上式结果，那么狭义遗传力就可以按一式计算

$$h^2 = \frac{V_A}{V_B} = \frac{S_a^2}{V_P} = \frac{1/2\,a^2}{V_P} = \frac{2[V_{F2} - \frac{1}{2}(V_{B1} - V_{B2})]}{V_{F2}}$$

式中的 V_{B1}，V_{B2} 和 V_{F2} 都可以通过直接测量来获得。

由于加性方差是基因纯合时与均值的偏差，而狭义遗传力又是加性方差和表型方差的比，也就是说加性效应在表型中产生多大作用。即基因纯合时遗传因素起了多大的作用。而育种中选种时只有选择纯合的类型才有意义，杂合的类型是不能作为种子的，所以狭义遗传力对种子的选育是有重要意义，狭义遗传力高的性状才可以被选育。

第四节　人类的数量性状

人类的大部分性状属于数量性状。例如皮肤的颜色，资料表明至少有四个基因在控制肤色，很可能是六个基因为主，带有 8～12 个加性等位基因和 9～13 个表型类别基因。更有趣的性状是智力和个性。我们若能搞清楚对于智力而言环境的因素（家庭的教养和社会环境的影响）和遗传因素（内在的自然因素）各起多大的作用将是很有益处的。若这些性状的遗传力是很高的，那么增加教育经费并不能使每一个人都成为人才。高智力和低智力的人可能需要不同类型的教育，智力较弱的人比那些智力强的人应能获得社会更多的帮助。若智力的遗传力是较低的（环境的影响较大），那么就需要花很多的钱来投入社会环境的改善。近年来的研究表明 IQ 的遗传力为 0.6，常在 0.8 左右，这是什么意思呢？遗传力是遗传方差和总方差的比。若对中等背景的孩子进行 IQ 测验，他们的亲属都接受过高等教育，他们都具有相同的机会，那么环境因素可能影响较低，而遗传力的影响是较高的。若整个的城市都包括在内，不论贫富，那么环境的方差是高的，因此遗传力就比较低。一个研究小组发现有一种特殊的等位基因（它是控制胰岛素样生长因子 2 受体）在天资较高的孩子中存在的频率较高（约 30%），而在一般的孩子中存在的频率为 16%。当然带有这种等位基因的每一个人都是正常的。

个性和行为必定含有遗传的成分，因为这些性状都要依赖于大量的神经递质及其受体。这就说明为什么还有心理药物。现在已分离了 100 多种神经递质，而且大部分都有受体，因此存在较高的遗传方差。血清紧张素（5－羟色胺）（serotonin）

受体基因的启动子有两个等位基因。一个具有高度翻译活性，另一个具有低度的翻译活性。带有高度活性等位基因的人比带有低度活性基因的人更易忧虑。双生子和同胞兄弟姐妹之间比较他们的“适应性”、“知觉性”、“外向性”、“神经过敏”和“直率”等项心理学测试结果表明豁达性格的遗传力为 30% ~ 45% （狭义遗传力为 10% ~ 25%）。这些遗传学的检测资料曾被纳粹用作统治工具，作为宣扬优等民族的借口，使人们对优生学和遗传素质产生误解。我们应当利用这些知识来提高人类的素质，使人们生活得更美好。

第七章

基因的结构和功能

基本概念

Beadle 和 Tatum 的“一基因一酶”学说具体表现了基因和酶之间的特殊关系。这个学说表明每个基因控制单个酶的合成或激活其活性。一个较新的说法是“一基因一多肽”。

先天性代谢缺陷也提示了酶与基因之间的关系

血红蛋白结构异常是一种分子病。分子病的发现表明基因的突变会导致蛋白质结构的改变。

噬菌体突变型的基因内作图产生此区域的精细作图，表明突变和重组的单位是 DNA 分子中的碱基对，基因内部有不同的突变位点，且可以发生重组。

顺反测验又称为互补测验，它是指将两个拟等位基因分别处于顺式或反式构型以观察表型是否有功能来判断它们是否属于同一顺反子内。若顺式有功能反式也有功能，则两个拟等位基因不在同一顺反子内；若顺式有功能，而反式无功能，则两个拟等位基因在同一顺反子内。

顺反子（cistron）是 Benzer 提出的一个概念，是指编码一个蛋白质的全部组成所需信息的最短片段，即一个基因。它是通过顺反测验所测定的遗传功能单位。经典的基因概念认为基因既是一个重组单位，突变单位又是一个功能单位。而顺反子的概念认为基因仅是一个功能单位，基因内部的碱基对才是重组单位和突变单位。

遗传学的研究始终围绕着一个中心，那就是基因的结构和功能。随着时间的推移，这方面的研究也在逐步地深入。摩尔根和他的弟子们通过艰苦卓越地工作将推测存在的遗传因子定位到了染色体上，并且呈线性排列。1927 年 Muller 用 X 射线引起了果蝇的突变，开始探索基因的结构和功能。在 40 年代初终于被正在摩尔根的麾下进行博士后研究的比德尔等所突破，因此而获得了诺贝尔奖。

第一节　一基因一酶学说

一、芽盘移植实验

1902 年英国的内科医生 A.Garrod 首先提出了酶与基因之间具有特殊关系的证据。当时他研究了四种遗传病（黑酸尿病，白化病，胱氨酸尿病和戊糖尿病），其中之一就是黑酸尿病（alcaptonuria），这是一种**先天性代谢的酶缺陷，属隐性**遗传病，患者由于黑尿酸氧化酶有缺陷，使代谢途经受阻，中间产物黑尿酶积聚在结缔组织内（图 7－1），有部分随尿排出，尿暴露在空气中成黑色。当时 Garrod 就指出，这些疾病是由于控制某一代谢步骤的酶活力的减退或缺失导致代谢阻滞的结果。他的发现和孟德尔的发现一样被埋没，直到 30 年代才重新发现他的学说。1952 年发现了一种糖原贮积症（Von Gierke 氏病），病人缺乏葡萄糖－6－磷酸酶，从而证实了 Garrod 的推论。有关酶和基因之间的关系的另一重要证据是 G.Beadle，和 B.Ephrssi 于 1935 年获得的。

斯特蒂文特（1919）报告了果蝇的朱砂眼突变型在一个雌雄嵌合体上回复成正常眼色，1933 年比德尔和伊弗鲁西（Ephrussi）从此报告中得到启发，认为这可能是研究基

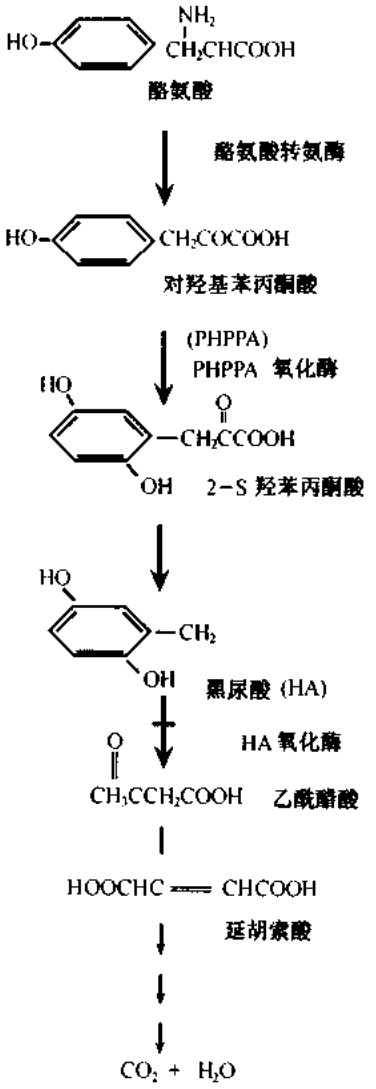


图 7-1　黑尿酸代谢途径。

因作用规律的一条线索。从而提出了采用成虫盘移植的方法来研究细胞内反应底物的改变对眼色的影响，但他们的课题申请未能得到洛克菲勒基金会的支持。作为加州理工学院生物学部主任的摩尔根敏锐地意识到这是一项具有挑战性的课题，毅然为他们提供了经费。在他们开展工作不久，一位研究昆虫变态的权威告诉他们俩人，这项移植实验是绝对行不通的。然而，他们却成功了。1937 年发表了题为“在果蝇眼色素合成的生化遗传中基因和酶的关系”的论文。

野生型果蝇的眼色是鲜红的（bright red），这是由棕色和桔黄色两种色素混合的结果。有 2 个涉及棕色色素合成的基因，这 2 个基因任何一个发生突变都不能产生棕色色素，而导致形成桔黄色眼(vermilion)，橙红色眼(cimnabar)，和暗红色眼(scarlet)。

首先比德尔和伊弗鲁西分离了两组在成体中可发育成眼的**芽盘**（imaginal discs）。所谓芽盘也叫**成虫盘**是全变态昆虫蛹期的外表皮内陷所长成的组织块，其中包含中胚层细胞。这些组织块将发育成成虫的各种器官，而幼虫的大部分器官则雕亡。当一个眼睛的成虫盘由一个幼虫取下移植到另一个幼虫的腹部，这个芽盘将发育成眼的结构，当幼虫发育成成虫时这种眼的结构可以其腹部分离出来。他们将野生型的成虫盘移到突变型果蝇幼虫的腹部来研究(图 7-2)，结果如表 7-1 所示。移植的野生型芽盘发育成鲜红色眼结构。移植的芽盘的表型和其本身的遗传信息相一致，并未受到环境的影响，这种不受周围细胞影响，按照自身基因型发育的现象称为**自主发育**（autonomous developnent）。当他们将桔黄眼（*V*）和橙红眼（*Cn*）突变体的芽盘移到正常果蝇的虫中，结果发育成正常的鲜红色眼结构(表 7-1，(2)，(4))。这种受到周围细胞影响，不按自身基因型发育的现象称**非自主发育**（nonantonomous development），他们推测这是由于野生型受体组织能为 *V* 和 *Cn* 突变型芽盘提供某些本身缺乏的物质，使其避免了遗传缺陷所造成的生化反应链的中断，从而合成了棕色色素。接着比德尔等又将 *V* 和 *Cn* 突变芽盘交互移植(表 7-1，(5)(6))，结果是桔黄眼（*V*）突变型芽盘在橙红色眼（*Cn*）突变型受体中能发育成正常鲜红色眼结构，而橙红眼（*Cn*）芽盘在桔黄眼果

表 7-1　眼成虫盘移植实验

	供体	受体	移植眼的颜色
(1)	+ *V*	*V*	+
(2)	*V*	+	+
(3)	+	*Cn*	+
(4)	*Cn*	+	+
(5)	*Cn*	*V*	橙红
(6)	*V*	*Cn*	+

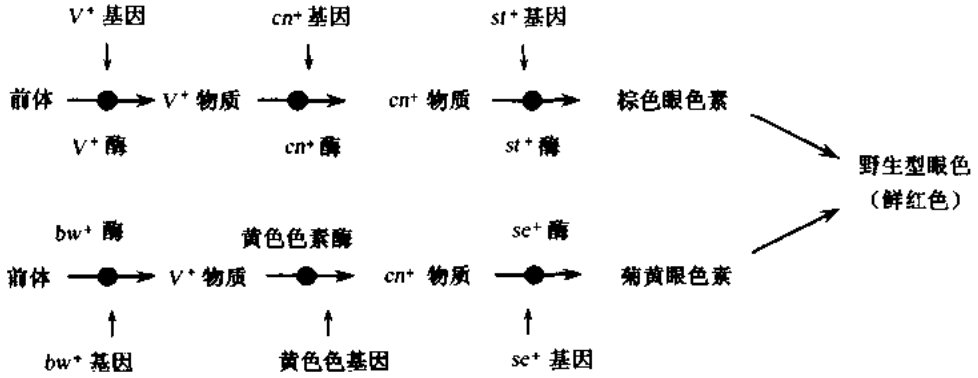


图 7-2 果蝇眼色红色色素合成的生化途径。
（参考 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 22-9）

蝇体内却不能发育成正常眼色的眼结构，而仍然和其基因型一致为橙红色。表明桔黄色突变型幼虫不能为 *Cn* 芽盘提供合成棕色色素的必要物质，看来这也是一种互补效应。

根据这个结果比德尔等推测棕色素的合成所涉及的生化反应键至少有两种中间产物，图 7-3 表明果蝇眼色素合成的生化步骤，及 V^+ 基因和 Cn^+ 基因的相互关系。*V* 突变型果蝇由于缺乏 V^+ 酶不能合成 V^+ 物质，所以反应中断，不能产生棕色色素，而另一条生化反应链是正常的，可以产生黄色的色素，结果果眼的眼色呈黄色。当将 *V* 突变型芽盘植入 *Cn* 突变型或野生型幼虫腹部时，周围的组织能提供其 V^+ 物质，而其 Cn^+ 是正常的，所以生化反应可以进行下去，直至合成棕色色素，和另一条链合成的桔黄色色素混合形成了鲜红色的野生型眼色；*Cn* 突变型果蝇由于 *Cn* 酶的缺陷不能合成 Cn^+ 物质，而有 V 物质的积聚，这种中间产物和桔黄色色素混和产生突变型橙红色眼色，当 *Cn* 芽盘植入野生型果蝇幼虫腹部时，周围的组织可以提供给它 Cn^+ 物质，使生化反应可以继续进行而产生鲜红的色素，但 *Cn* 芽盘移植到 *V* 突变型幼虫腹时，周围的组织因 V 酶的缺乏，无法提供它所缺乏的 Cn^+ 物质，所以结果眼色仍呈橙红色。以上表明一种基因控制着代谢的各个步骤，而每一步骤又都被特定的酶所控制，比德尔从巴黎回国时，经过斯坦福大学，在那儿，他和 Tatum （塔特姆）合作，用化学方法鉴定与眼睛颜色相关的物质，但是失败了，而现在人们早已搞清了这一代谢途径的化学物质和酶（图 7-3）。

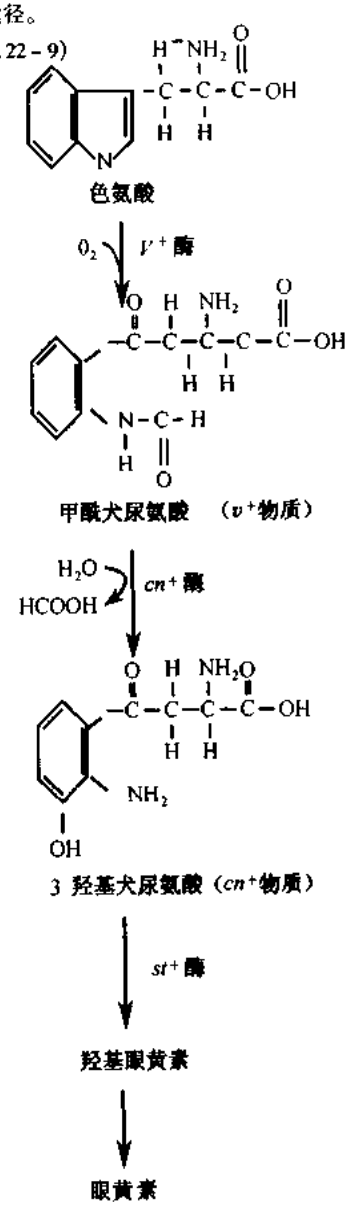


图 7-3 果蝇眼色的棕色素合成生化反应途径。

二、一基因一酶学说

比德尔的芽盘移植实验已表明酶的特异性是在基因

水平上来控制的，使基因和酶的关系已达到了“呼之欲出”的地步，但在当时的条件下，他们遇到了一个巨大的困难，那就是眼睛颜色相关物质鉴定，因此无法进一步确定基因和酶的关系。他们已意识到当时研究方法存在着一定的局限性：①研究者一般只局限于研究非致死性的遗传性状，这些性状在某种程度上与非主要反应有关；②研究所选用的性状必须具有明显的表型。这样的性状大都和形态变异有关，与他们有关的生化反应系统似乎都是相当复杂，很难分析的。此时摩尔根的研究室已采用了红色链胞霉（*Neurospora Crassa*）作为实验材料。比德尔和塔特姆认为这是一种理想的研究材料，因为①其遗传特性已了解清楚；②用射线或紫外照射孢子，很容易诱发突变；③可以在特定的培养基上生长，并能人工合成它所必需的全部物质。1940 年他们又创立了一种研究代谢途径遗传控制的新方法，即用 X 射线处理诱发那些控制已知特异生化反应的基因，使之发生突变。如果菌体必须某种生化反应才能在一种特定的培养基上生存下来，那么一个不能完成这种反应的**突变体**在这种培养基上显然会死亡。但若在这种培养基中加入这个反应必需的主要产物，突变体就能成活的话，突变体就能保存下来，并用于研究。分离突变型的实验程序如图 7－4 所示，把一个个单个的红色链胞霉的孢子接种于含完全培养基（野生型和各种缺陷型都能生长）的多个试管中，然后把在完全培养基上长出的真菌转接到基本培养基上（仅野生型才能生长），他们发现有 1/200 的孢子只能在完全培养基上生长，从而推测这些孢子是突变子代。在这些孢子中可能由于一个基因的突变而阻断了一条基本代谢途径。突变体可以在完全培养基上生长，那是由于完全培养基提供了代谢所需要但本身又不能合成的物质。因此，为了识别不同的缺陷，就把突变体接种到一系列补加了各种氨基酸，维生素，嘌呤或嘧啶等的基本培养基上，若能生长，说明这种突变所缺乏的物质正是基本培养基上所补加的物质。他们用 X 射线诱导产生 380 种突变型，然后将它们进行杂交，对所产生的 6800 个以上的**子囊孢子**逐个进行了检查。这项研究终于得出了这样的结论：基因对控制特定的生化反应的酶的合成在起主要作用，即“一基因—酶（one gene，one enyme）”学说。

现举例说明精氨酸合成代谢的一系列突变品系。首先人们发现四个突变位点分居于不同的染色体上。分别称为 *arg*E，*arg*F 和 *arg*G 和 *arg*H。人们发现这四个位点每种**缺陷型**对鸟氨酸、瓜氨酸、精氨琥珀酸和精氨酸的反应是不同的（表 7－2），野生型菌株在基本培养基上可以生长，在基本培养基中不论加四种氨基酸中的任何一种，*arg*F 突变品系都能生长；*arg*E 品系在加了鸟氨酸的培养基上仍不能生长，而加其余 3 种氨酸都能生长，*arg*G 品系在加了精氨琥珀酸或精氨酸时都能生长；而 *arg*H 品系仅生长于加了精氨酸的培养基上。

表 7－2　精氨酸营养缺陷型的生长反应

菌株	基本培养基加不同的氨基酸				
	不加	鸟氨酸	瓜氨酸	精氨琥珀酸	精氨酸
野生型	+	+	+	+	+
arg E	-	+	+	+	+
arg F	-	-	+	+	+
arg G	-	-	-	+	+
arg H	-	-	-	-	+

“+”表示能生长，“-”表示不能生长

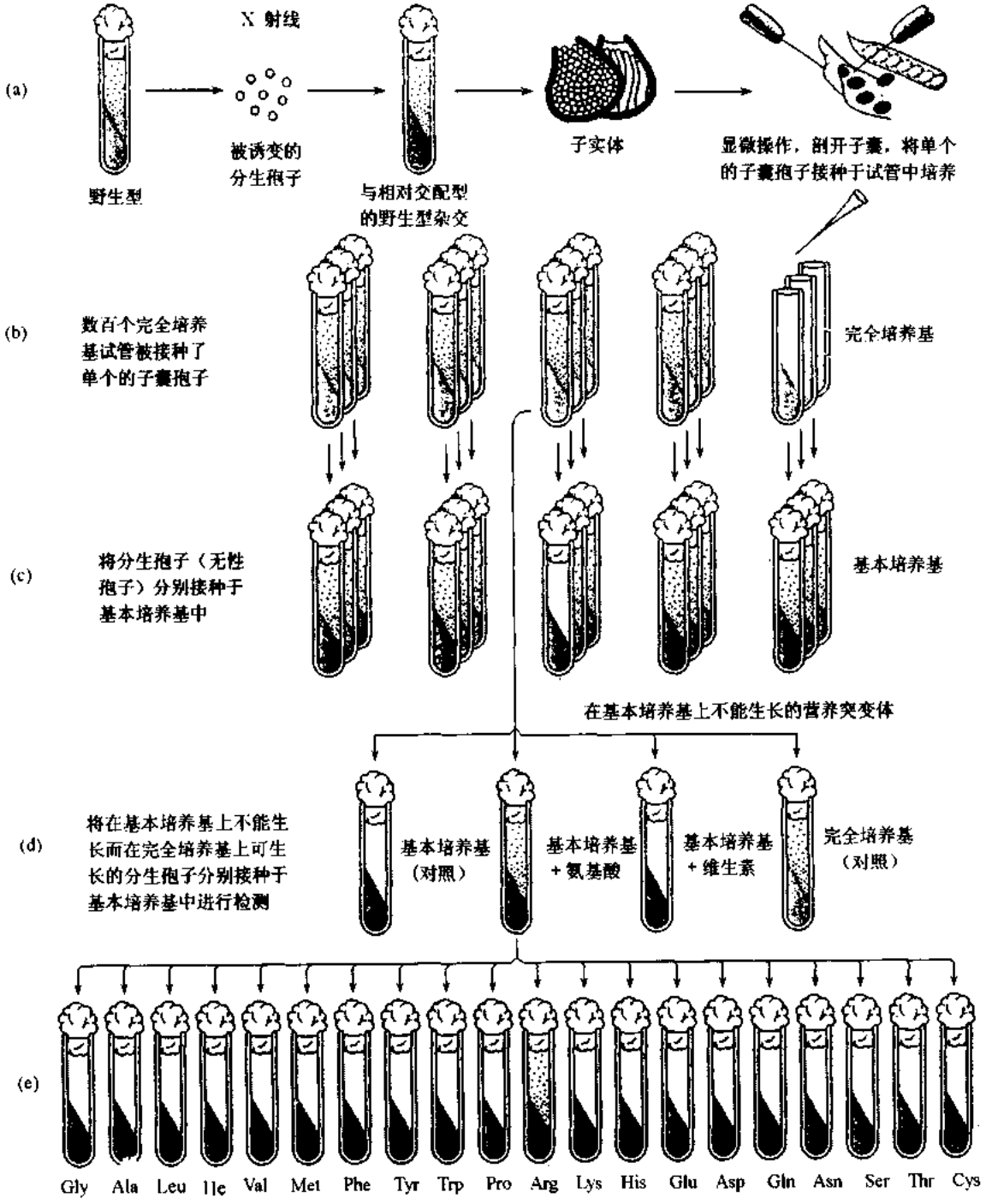


图 7－4　分离突变型子囊孢子的实验过程。

（转引自 Griffiths A.J.F. *et al*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 12－1）

这个实验说明 *arg*E 品系是乙酰鸟氨酸酶的缺陷，从而不能合成鸟氨酸，使代谢键受阻，其他步骤的酶都是正常的，所以只要加鸟氨酸或鸟氨酸后面的任何一种产物，都可以使生化链得以完成，缺陷型链孢霉就得以成活，同理其他的缺陷品系也是表明了具有相应的酶缺陷。实验的结果和已知的精氨酸生物合成途径（图 7－5）的顺序完全

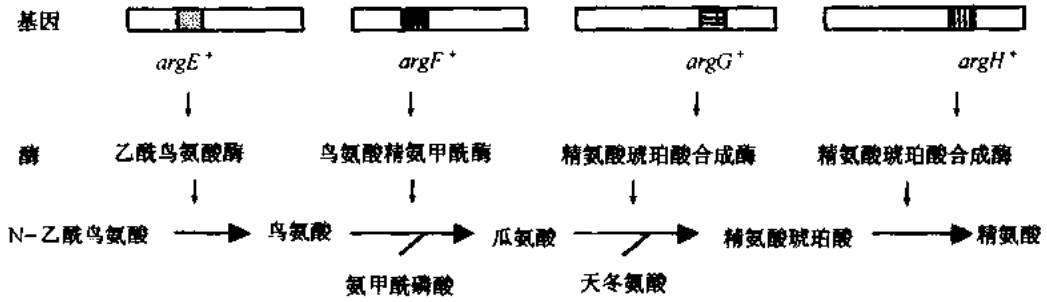


图 7-5　精氨酸生物合成途径(参考 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 22-9)

相符，很好地表明了一个基因控制一个酶的学说。一基因一酶学说首先阐明了基因是通过对酶的控制来决定性状的，对以后基因的结构和功能的研究产生了深刻的影响，对分子遗传学的诞生起了重要的助产的作用。因此此项研究于 1958 年获得诺贝尔奖。但这一学说仍处在遗传学研究的初期，属于"经典遗传学"范畴，基因的概念仍是一个突变单位，重组单位和功能单位。这个学说也还不能说明基因如何控制酶的合成。这是由于历史的局限。现在看来"一基因一酶学说"基本上仍是正确的，只不过人们已知道基因的功能是编码蛋白质，而蛋白质的功能是多样的，有结构蛋白，转运蛋白，免疫蛋白，受体蛋白等等，不限于酶这种形式，所以也有人认为这个学说称为"一基因一蛋白"更为合适；随着科学发展，人们又了解到有的蛋白并不是由一个基因控制合成的。如人类免疫球蛋白由两条重链和两条轻链组成，重链是由 14 号染色体上的基因编码，而轻链中的 κ 链是由 2 号染色体编码，λ 链是由 22 号染色体上的基因编码的，因此有人提出以"一基因一多肽"这一概念，无论什么提法，其精髓是不变的。

第二节　人类酶缺陷的遗传基础

人类的**遗传性酶病**（hereditary enzymopathy）或称**遗传性代谢病**（inherited metabolic disease）是指由于单个基因突变导致酶的结构与功能的改变或量的改变，从而引起了所催化的生化反应链的改变所产生的一类疾病(表 7-3)。据 Mckusick 1988 年的统计这种遗传性代谢病共 232 种。先天性酶缺陷也表明了"一基因一酶"的原理。通过对先天性酶缺陷的研究也为基因编码进一步地提供了证据。

表 7-3　人类中部分酶缺陷遗传病

	病名	酶缺陷
Acid phophatase deficiency	酸性磷酸酶缺乏症	酸性磷酸酶
Albinism	白化	酪氨酸酶
Alkaptonuria	黑尿酸症	黑尿酸氧化酶
Ataxia, infermittenta	间歇性共济失调	丙酮酸脱羧酶
Disaccharide intolerance	二糖无耐受力	蔗糖酶
Fructose intolerance	果糖无耐受力	果糖-1-磷醛缩酶
Fructosuria	果糖尿	肝果糖激酶
G6PD deficiency(favism)a	葡萄糖-6-P脱氢酶缺乏症(蚕豆病)	葡萄糖-6-P-脱氢酶
Glycogen storage diseasea	糖原贮积病1	葡萄糖-6-P-脱氢酶

续表

	病名	酶缺陷
Gout, primary	痛风	次黄嘌呤磷酸核糖转移酶
Hemolytic anemia	溶血性贫血	谷胱甘肽过氧化物酶
Hemolytic anemia	溶血性贫血	乙糖激酶
Hemolytic anemia	溶血性贫血	丙酮酸激酶
Hypoglycemia and acidosis	低血糖和酶中毒	果糖－1,6－二磷酸酶
Immunodeficiencya	免疫缺乏症	腺苷脱氨酶
Immunodeficiency	免疫缺乏症	嘌呤核苷磷酸化酶
Immunodeficiency	免疫缺乏症	尿苷－磷酸激酶
Intestinal lactase deficiency (adult)	肠乳糖酶缺乏症(成人型)	乳糖酶
ketoacidosisa	酮酸中毒	琥珀酰COA:3－酮酸辅酶A转移酶
Leigh's necrotizing Encephalomelopathya	坏死性脑脊髓病	丙酮酸羧化酶
Lesch－Nyhan syndromea	莱－尼氏综合征(自毁容貌综合症)	次黄嘌呤磷酸核糖转移酶
Lysine intolerance	赖氨酸不耐症	赖氨酸:辅酶Ⅰ氧化还原酶
Male pseudohermaphroditism	男性假两性畸形	睾丸17,20－碳链裂解酶
Maple sugar urine diseasea	枫糖尿病(Ⅰ型)	酮酸脱羧酶
Orotic aciduria	乳清酸尿症	乳清苷酸脱羧酶
Phenylketonuria	苯丙酮尿症	苯丙酮酸羧酶
Porplyria, acutea	急性卟啉a症	尿卟啉原－Ⅰ合成酶
Porphyria, congenital erythropoietic	先天性血性卟啉症	尿卟啉原－Ⅲ合成酶
Pulmonary emphysema	肺气肿	α－1－抗胰蛋白酶
Pyridoxine－dependent infantile convulsions	依赖于吡哆醇婴儿惊厥	谷氨酸脱羧基酶
Pyridoxine－responsive anemia	吡哆醇反应性贫血	λ－氨基乙酰丙酸合成酶
Kidney tubular axidosis with deafness	???的肾小管酸中毒	碳酸苷酶
Rickett, vitamin D－dependent	维生素D依赖性佝偻病	25－羟胆钙化醇（羟维生素D_3）－羟化酶
Tay-Sachs disease	泰－萨二氏病(家族性？？性白痴)	乙糖酰胺酶A
Thyroid hormone synthesis, defect in	甲状腺激素合成缺陷	碘化物过氧化物酶
Thyroid hormone synthesis, defect in	甲状腺激素合成缺陷	脱碘酶
Tyrosinemia	酪氨酸血症	对羟基苯丙酮酸氧化酶
Xeroderma pigmentosum	着色性干皮病	DNA－特异核酸内切酶

一、半乳糖血症

半乳糖血症（galactosemia）是婴儿患的一种先天性代谢病，**属常染色体隐性遗传**，基因频率为32‰，群体发病率为1/10万。正常情况下，婴儿哺乳后，乳中所含的乳糖经消化酶分解为半乳糖和葡萄糖，半乳糖在激酶的作用下形成半乳糖－1－磷酸，进而在肝脏中经半乳糖－1－磷酸尿苷转移酶（galactose-1-phosthate uridyltransrerase Gal-1-PUT）的催化转变成葡萄糖－1－磷酸，进入**酵解途径**。若由于基因突变引起了Gal-1-PUT结构改变而失去功能，那么半乳糖代谢阻断，半乳糖积聚在血及组织内，并随尿排出。中间产物半乳糖－1－磷酸（Gal-1-P）对细胞有害，主要侵犯肝，肾，脑和晶状体，故患儿

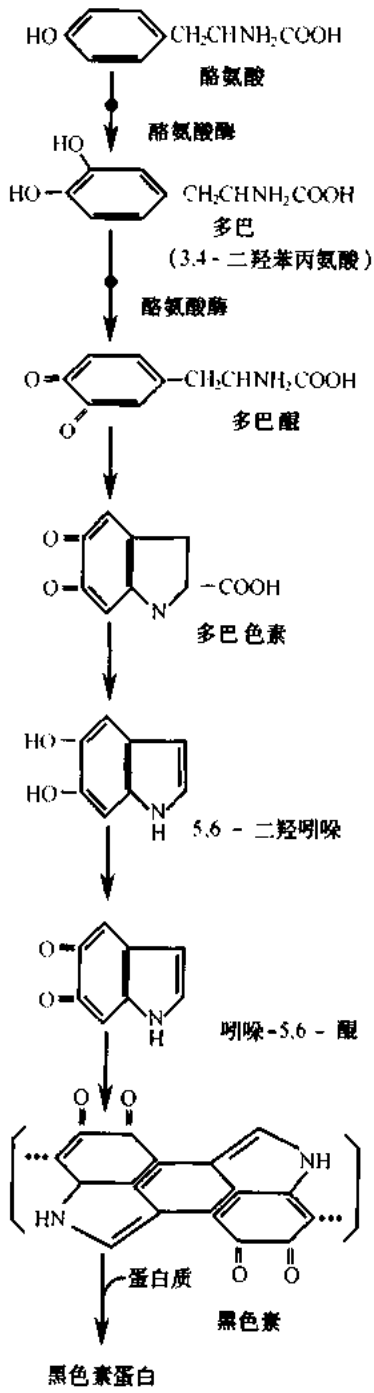


图 7-6　黑色素合成的生化途径。

哺乳后数日出现呕吐、腹泻、脱水等症状。一周后，肝脏肿大，出现黄疸、腹水和白内障。数月后出现智力发育障碍，常夭折。这就是半乳糖血症。若出生后不喂以乳类和乳制品，婴儿能完全正常地发育。若中途停止乳类食物可改善症状，但智力不能恢复。

二、白化症

白化病(albinism)是 Garrod 最初研究的四种遗传代谢病之一 。属常染色体隐性遗传。发病率为 1/17000。白化病是酪氨酸代中的一种先天性酶缺陷，不能形成黑色素。不仅在人类中有白化，其他哺乳动物、鸟类、两栖、爬行甚至鱼类中都有同类型的突变，白猿，白虎，白蛇，白麻雀，白鱼都时有报道。

白化患者全身黑色素细胞均缺乏酪氨酸酶，不能使酪氨酸变成黑色素，皮肤呈白色或浅红色，毛发银白或浅黄色，虹膜和脉络膜都不含色素，因而虹膜红色，并且透明。

黑色素在**黑色素体**(melanosome)里合成，黑色素体位于黑色素细胞中，这种细胞是一种特殊的树枝状细胞，分布于皮肤真皮和表交界处的毛球。皮肤的颜色不取决于黑色素细胞的多少而仅仅依赖于黑色素产生的量。黑色素体由黑色素细胞树状突进入周围的**角化细胞**(Keratinocyte)中，覆盖在细胞核周围，使其核免受阳光中紫外线的损伤。

黑色素的生化合成较为复杂，整个的生化反应都在黑色素体内进行，反应中有 2 个步骤都需要酪氨酸酶催化（图 7-6）酪氨酸酶由黑色素细胞的核糖体合成，以一种可溶的形式通过内质网传递给高尔基体，再转移到黑色素体，随着黑色素体的发育而催化黑色素的合成。一旦酪氨酸酶有先天缺陷，就使得黑色素合成途径受阻而导致白化。

三、苯丙酮尿症

苯丙酮尿症(phenylketonuria，PKU)于 1934 年首次被发现，是研究得最透彻和治疗效果最好的一种氨基酸代谢缺陷。我国的发病率为 1/16500；白人发病率为 1/万。患者皮肤、毛发色浅，汗和尿有特殊的鼠嗅味，智力低下，缺乏表情，易激动，有的常发生惊厥，多于 30 岁前死亡。原先发现的 PKU 称为**典型 PKU**(高苯丙氨酸血症Ⅰ型)，后又研究了 PKU 的变型或其他的苯丙氨酸代谢缺陷。

典型 PKU 由于肝脏中缺乏苯丙氨酸羟化酶，来自食物中的苯氨酸不能形成酪氨酸，而经转氨酶的作用转变为苯丙酮酸。大量苯丙氨酸及苯丙酮酸聚积在血和脑脊液中并随尿排出。由于血液及体液中苯丙氨酸过剩而竞争性抑制酪氨酸酶，酪氨酸转化为多巴的反应受阻，因而影响到黑色素的合成，所以患者皮肤和毛发含黑色素少，颜色较浅。多巴又是肾上腺素合成的中间物，酪氨酸代谢障碍使体内肾上腺素浓度降低(图 7－7)。

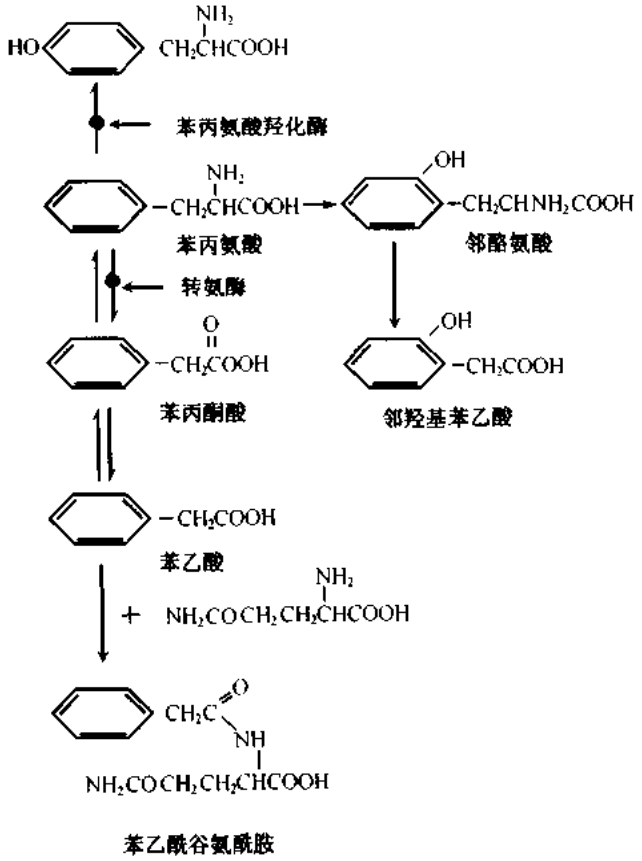


图 7－7　苯丙酮尿症的生化紊乱。

大量的苯丙氨酸除分解为邻羟基苯乙酸外，还通过苯丙酮酸代谢产生苯乳酸，苯乙酸和苯乙酰谷酰胺，大量从尿中或汗中排出，发出鼠嗅味。

患者的智力低下和各种神经学异常可能因体内大量积聚苯丙氨酸，竞争抑制了酪氨酸和其他氨基酸转运到大脑，影响了大脑的发育；也可能由于苯丙氨酸的异常代谢产物。抑制脑组织 L 谷氨酸脱羧酶的活性，使 γ－氨基丁酸生成量减少，而损伤脑细胞和生理功能。

在一些发达国家，可生产不含苯丙氨酸的食物，供患者食用。若早期诊断并食用不含苯丙氨酸的食物，患儿可正常发育。

四、莱－尼二氏综合征

莱－尼二氏综合征(Lesch-Nyhan syndrome)又称为自毁容貌综合征。属 X－连锁隐性遗传仅发现于男性。发病率为 1/30 万。患儿出生正常，8 个月到 1 岁时出现肌张力

障碍，指痉病和舞蹈症。患儿 2～3 岁开始强制性地吃自己的手指、嘴唇和口腔粘膜，并侵害他人，必须捆绑其双手或拔掉牙齿。故称为自毁容貌综合征。患儿还表现为高尿酸血症，约在 12 岁左右夭折。

本病患者由于缺乏次黄嘌呤鸟嘌呤磷酸核苷转移酶（HGPRT），因此合成肌苷酸（IMP）和鸟苷酸（GMP）的量不足，则不能维持神经系统的正常功能，特别是基底神经节的正常功能，从而导致了神经功能的异常。HGPRT 的缺乏同时使中枢神经系统中次黄嘌呤和黄嘌呤合成过多（图 7－8）而产生毒害效应，在患者的脑脊液中次黄嘌呤浓度

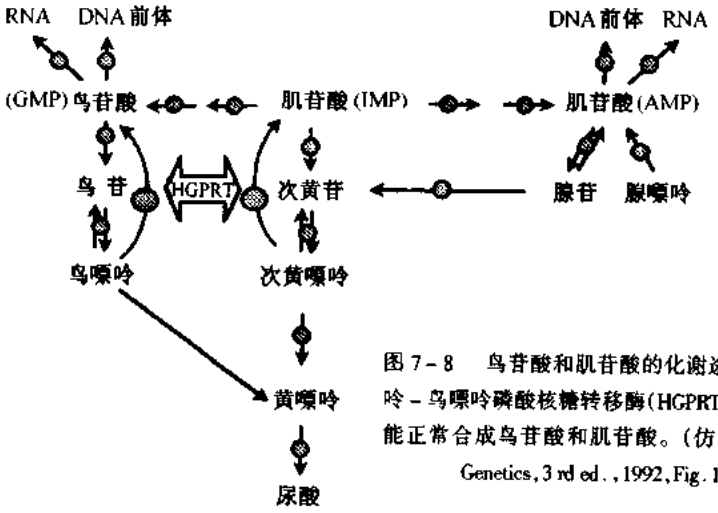


图 7－8　鸟苷酸和肌苷酸的化谢途径。当次黄嘌呤－鸟嘌呤磷酸核糖转移酶（HGPRT）有缺陷时，不能正常合成鸟苷酸和肌苷酸。（仿 Russell，P.J.：Genetics，3 rd ed.，1992，Fig. 16－13）

为正常人的 4 倍，而动物实验证明，用甲基黄嘌呤喂养大鼠可使其产生类似Lesch-Nyhan 综合征的自毁损伤的神经学症状。由于患者体内嘌呤的生物合成过多，从而使尿酸合成过量，产生高尿酸血症。

采用放射自显影技术证明培养的女性杂合体成纤维细胞中，60% 的 HGPRT 有活力，余下 40% 没有活力，符合 X－连锁遗传的 Lyon 假说。

第三节　遗传咨询和产前诊断

一、遗传咨询

遗传咨询（genetic counseling）是专家回答咨询者的一切问题，并提供所涉及遗传缺陷的诊断，遗传方式，再发风险，防治措施，预后以及是否适宜生育等信息，由咨询者根据意见，权衡利弊，作出各种抉择。

咨询时根据**先征者**的完整**病史**，**家族史**及临床检查资料来诊断是否是遗传病，并分析推算再发风险，即家族中未来孩子患有这种遗传的概率。必要时必须补做某些临床检验，携带者检测。由于遗传病有四千多种，不易全面掌握，必要时可查阅 McKusick V.A 所著的 *Mendlian Inheritance in Man* 来帮助分析。

若高龄孕妇担心胎儿由于染色体畸变产生畸形染色体病，或者有家族遗传病史的孕妇想知道早期胎儿是否有遗传缺陷，有的可能通过产前诊断得到准确的信息，以便决定是否须要终止妊娠。

二、产前诊断

产前诊断（prenatal diagnosis）是对早期胎儿是否患有遗传缺陷或先天畸形作出准确的判断。产前诊断的指征是：①亲体有染色体异常或先天性代谢异常以及分子病的携带者；②严重伴性遗传的基因携带者；③母亲曾生育过先天愚型（21－三体）或其他先天性异常的婴儿；④35～40岁以上的高龄孕妇；⑤过去曾有多次不明原因的流产，早产和死胎的孕妇；⑥亲体（特别是早孕期间的母亲）曾多次接受过放射线的照射或在妊娠早期服过一些可以引起胎儿畸形的药物。

由于羊水中具有胎儿的脱落细胞，通过分离培养、DNA的分析以及核型分析，就可以检测出某些基因突变及染色体畸变。通过X和Y染色体的鉴别可以确定胎儿的性别，这对严重的X－隐性遗传缺陷的家族来说也是很有意义的。例如自毁容貌综合征，进行性肌营养不良（杜兴氏症）都是男婴患病，且可致死。若检出是男胎就可终止妊娠，以免冒50%的可能生育患儿的风险。另外根据酶的活性及量的测定，羊水中代谢产物的变化可以诊断出某些先天性代谢缺陷。

产前诊断采集标本的方法通常是通过**羊膜穿刺**抽取羊水，收集胎儿脱落细胞供产前诊断（图7－9），也可以从孕妇的宫颈粘液中获取**绒毛膜细胞**，或者直接从子宫腔中吸取绒毛膜细胞进行检测（图7－10）。

超声波可用于产前诊断神经管缺陷，如无脑儿，脊椎裂和水脑儿。若辅以测定羊水的甲胎蛋白（αFP）浓度和乙酰胆碱酯酶（AchE）的活性可以提高确诊率。还可以检测胎儿发育异常，胎儿性别，以及母胎血型不合而引起的胎儿胸腹水及全身水肿，胎盘肿大等。

胎儿镜（fetoscope）又称**羊膜镜**或**宫腔镜**，可以从宫颈口插入妊娠14～18图的子宫腔内及羊膜腔内直接观察胎儿，辨认形态上的畸形，识别性别，还可以取活组织和血液标本。胎儿中有的蛋白和酶仅存在于血液中，目前血液中蛋白和酶的先天性缺陷也不是都能用DNA的各种检测来确定。所以用羊水和羊水细胞都不能对这类缺陷进行产前诊断，而用胎儿镜就可弥补这方面的不足。

三、携带者的检出

携带者（carrier）一般是指表型正常，但带有有害的基因的杂合体和具有染色体平衡易位的个体。携带者都可将有害基因传递给后代，是群体中使有害基因的频率保持稳定的重要因素。因此检出杂合体对于预防有害基因的纯合，降低群体中有害基因的频率都是十分重要的。若查出的携带者尚未结婚则可建议及劝阻他们不要和另一同类型携带者结婚。若偶尔两个同类型携带者已经结婚，那么就要建议他们不要生育，或妊娠时做产前诊断。在意大利，在地中海贫血的携带者频率高达10%，所以法律规定所有儿童均须接受相应的携带者检查，携带者之间禁止相互结婚，从而有效地降低了患者的发生率。神经节苷酯贮积病GM2－I型又称为Tay－Sachs症，或**家族性黑朦性痴呆**（familial amaurotic idilcy），或氨基己糖苷酶A（hexosaminidase A，Hex A）缺乏症。属隐性遗传患儿出生后3－6个月出现症状，神情淡漠，不能抬头，痴呆，渐出现癫痫样发作，视力减退以至失明。显著消瘦，肌张力减退，四肢瘫痪，3岁左右夭亡。本病由于溶酶体氨基己糖苷酶A缺乏，使神经苷脂GM2贮积（图7－11）所致。犹太人中发病较多。美国犹太人中，本病基因频率为0.013－0.016，杂合子占3.24%，酶基因定位于15q23－

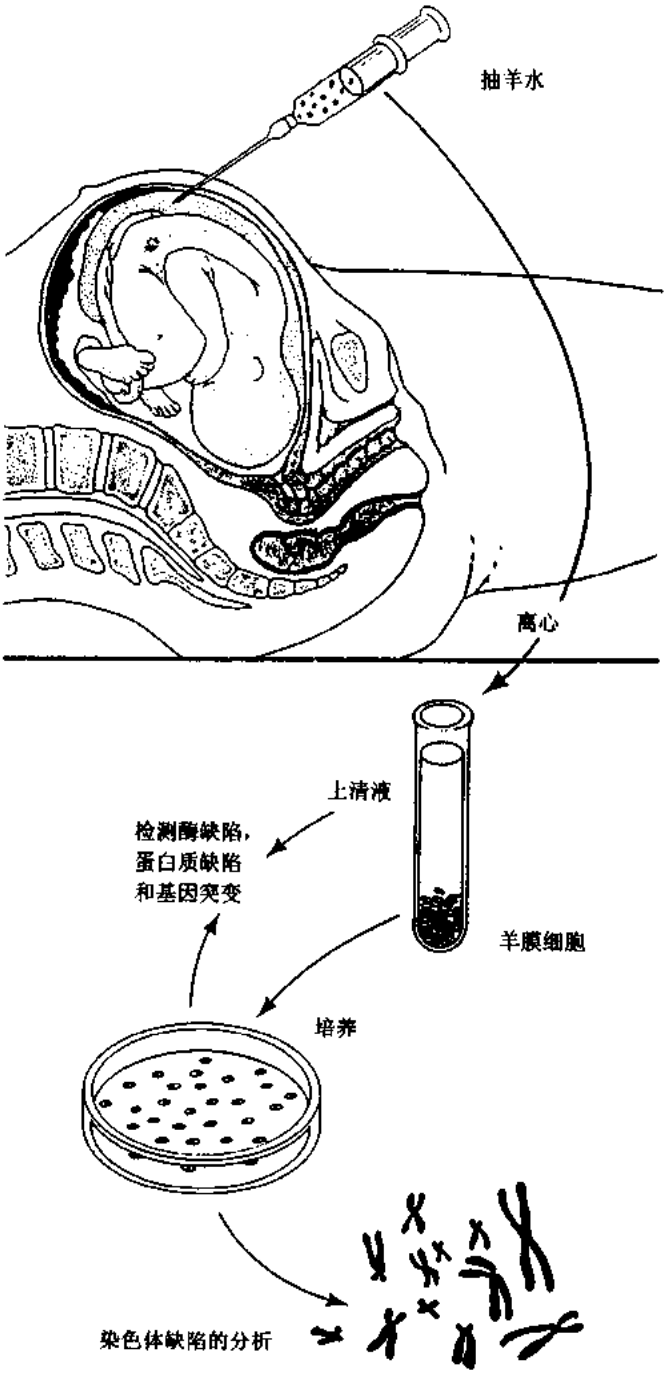


图 7－9　用于遗传病产前诊断的羊膜穿刺术的步骤。
（仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 8－16）

q24。自 1969 年起，13 个国家的 73 个城市开展了大规模的携带者筛查，使本病的发生率降低 65% ~ 85%，成为防止遗传病的范例。

筛查携带者的方法有以下几种：

(1) 染色体核型分析，以筛查带有染色体平衡易位者。为明确诊断可用 G 分带的方法确定易位。

(2)酶活性的检测。用这种方法的前提是在检测时，基因已经表达。有的基因到50

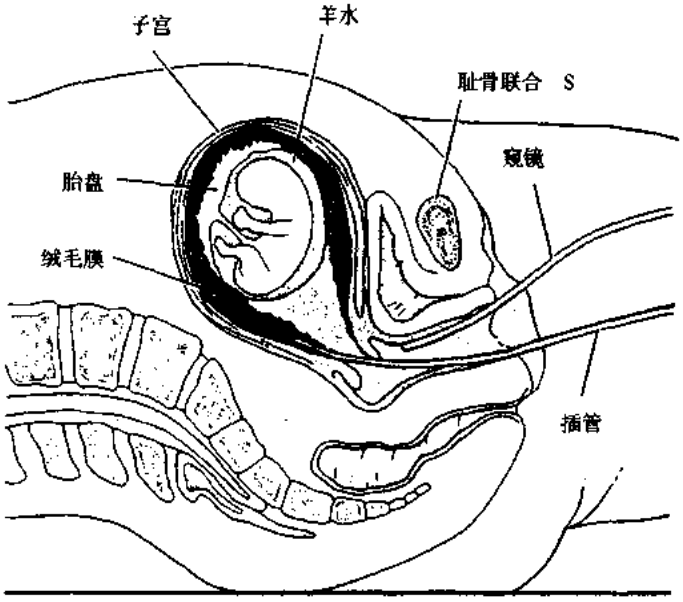


图 7－10　直接从宫腔中吸取绒毛膜细胞进行检测一种早期产前诊断遗传缺陷的方法。
（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 8－17）

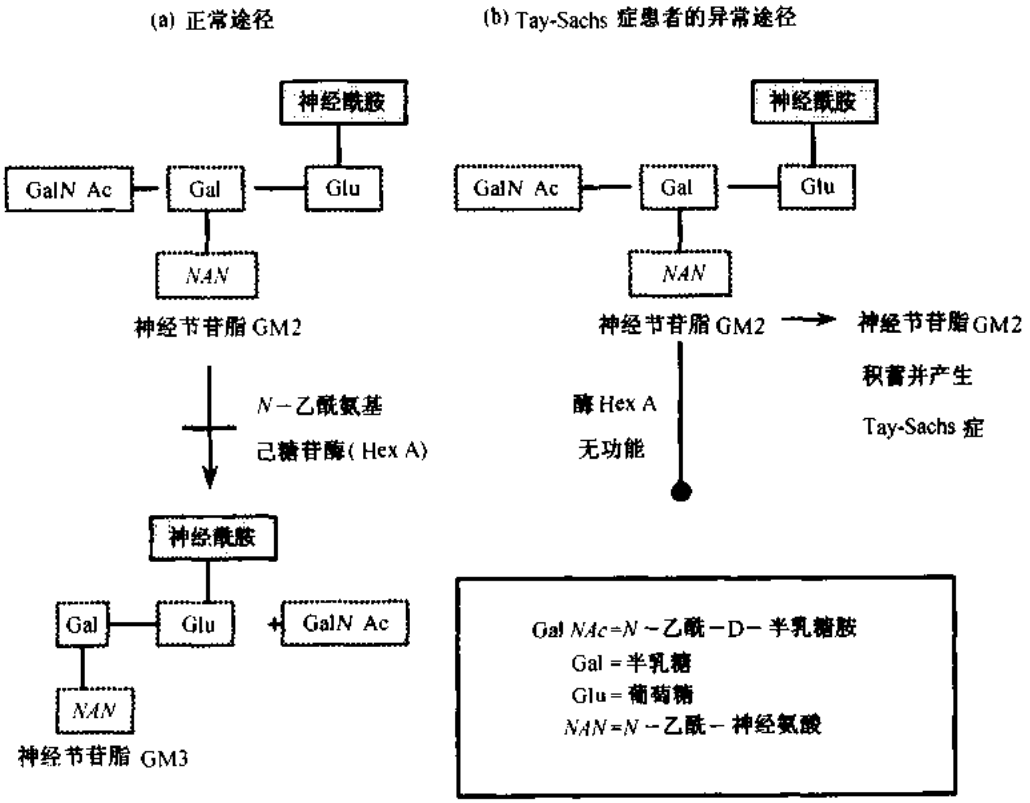


图 7－11　神经节苷脂 GM2 贮积病。（仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 8－14）

岁左右才表达，那么在青少年中就不宜用此法。另外有的酶和其产物具有明显的剂量效应，也就是说基因的拷贝数和产物成正相关则是可以采用此法。而有的基因只需其

中一个等位基因是正常的就可以产生足够的酶来完成生化反应，那么就不能用此法。如苯酮尿症的携带者其肝中苯丙氨酸羟化酶的活性只有正常人的一半；半乳糖血症的携带其红细胞中半乳糖－1－磷酸尿苷转移酶的活性也是正常人的一半。黑朦性痴呆的携带者 bexA 的活性比正常人的明显降低。

(3)分子生物的方法。一些不宜用生化方法检测携带者的可采用分子生物学的方法若突变是由缺失所造成，那么可用特殊的探针进行分子杂交，结合的探针剂量可反映出基因的拷贝数。若突变位点正好在酶切位点上，那么也可以用限制性酶切割某基因后进行电泳，通过 DNA 片段的多少和大小就可以加以确定，如镰刀形贫血(见第 20 章)。还可以用限制性酶切片段长度的多态性加以筛选(见第 10 章)。

四、分子病

分子病(molecular disease)这一概念是由 Pauling 其同事于 1949 年提出的。分子病是指由于基因的突变引起了蛋白质分子结构的改变而导致的疾病，如镰形细胞贫血(sickle cell anenia)。

镰刀形贫血是 J.Herrick 首次报道。他发现患者的红细胞在缺氧时不是球形的，而呈镰刀形。镰刀形细胞比正常细胞要脆弱得多，容易破裂。1949 年 Pauling 等通过正常人和镰刀形贫血患者之间血红蛋白的自由泳动率差异分析，证明了镰刀形贫血是一种由于蛋白质结构改变而引起的分子病。他们从正常人，患者及杂合体中分离了血红蛋白，然后进行电泳，结果如图 7－12 所示。正常人的血红蛋白(Hb－A)只有一条带，移动较慢；患者的血红蛋白(Hb－S)也只有一条带，移动较快；而杂合体的血红蛋白出现 2 条带，迁移率分别和 Hb－A，Hb－S 相同，他们认为镰刀形贫血是由于突变引起了血红蛋白结构改变的结果。这个实验第一次证明蛋白质的结构是由单个基因控制的。

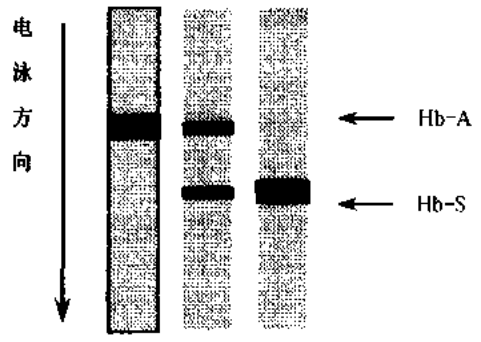


图 7－12　正常血红蛋白 $\beta^A\beta^A$、镰型贫血的血红蛋白 $\beta^S\beta^S$ 以及杂合血红蛋白 $\beta^A\beta^S$ 的电泳图谱。(仿 Russell, P.J.: Genetics, 3rd ed., 1992, Fig.8－23)

1957 年 V.M.Ingram 又证明了镰刀形贫血患者仅仅是单个氨基酸替换（图 7－13)。

人类的血红蛋白由**珠蛋白**和**血红素**(neme)所构成。珠蛋白共有四条肽链，二条 α 链，二条 β 链。α 链有 141 个氨基组成，β 链有 146 个氨基酸组成。每一条肽链都和 1 个扁平状的血红素结合，构成一个血红蛋白单体。

现在已知 α 链是由位于 16 号染色体上的 α 基因控制合成；而 β 键是由位于 11 号染色体上的 β 基因控制合成。由于 α 和 β 链基因的缺陷会导致 α 链的缺乏产生了血红蛋白病，分别称为 **α 地中海贫血**和 β 地中海贫血。此类血红蛋白病是由于血红蛋白合成数量的异常，而另一类是由于基因的突变引起了血红蛋白结构的异常。在人类中现已发现 13 种异常，镰刀形贫血是其中一种，它的 β 链第 6 位的谷氨酸被缬氨酸替代(图 7－13)，纯合的 Hb*s* 血中的 Hbs 蛋白含量达 90%以上，出现镰刀形贫血，大多幼年夭亡。杂合体含 Hbs 含蛋白约为 20%～40%，一般没有临床症状，称为**镰刀形细**

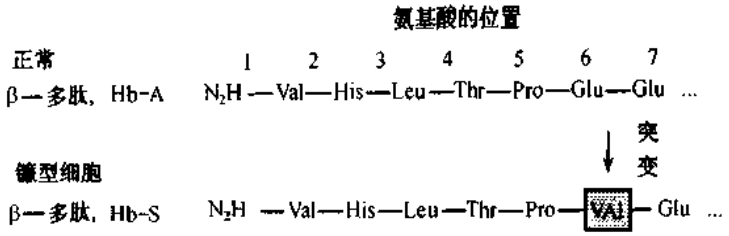


图 7－13　镰型贫血中单个氨基酸的改变。(仿 Russell, P.J.: Genetics, 3 rd ed., 1992, Fig. 8－24)

胞特征(sickle cell trait)。镰刀形贫血由于亲水性谷氨酸(带有负电荷和极性侧链)被疏水性的缬氨酸（不带电荷，有非极性侧链）所取代，使 Hbs 在脱氧状态下，比正常的 HbA 的溶解度低 5 倍，于是在氧张力低的毛细血管区，Hbs 溶解度大大降低而形成管状结构和凝胶化，导致红细胞扭曲成镰刀状，称为镰变。镰变是可逆的，氧张力增加时，红细胞能够恢复。但如果镰变时间延长，细胞膜便遭到不可逆的损害，结果这种镰状细胞被网状内皮细胞从循环系统中排除，出现慢性溶血性贫血症状。

其他位置上的氨基酸的改变也可引起不同的血红蛋白贫血（图 7－14）但并不都像镰形贫血那样严重。如 Hb－C 的纯合体也仅表现为轻度贫血。我们从图中可以看出：①一个氨基酸的改变就有可能影响到整个蛋白质的功能；②血红蛋白的异常可能是由不同位置上的氨基酸改变所引起。这就意味基因的突变可以发生在内部不同的位点上(图 7－14)，而不是整个的基因发生改变。

	序号	1	16	30	57	58	68	141
α-链	正常	Val	Lys	Glu	Gly	His	Asn	Arg
	Hb-I	Val	Asp	Glu	Gly	His	Asn	Arg
	Hb-G Honolulu	Val	Lys	Gln	Gly	His	Asn	Arg
	Hb Norfolk	Val	Lys	Glu	Asp	Gly	Asn	Arg
	Hb-M Boston	Val	Lys	Glu	Gly	Tyr	Asn	Arg
	Hb-G Philadelphia	Val	Lys	Glu	Gly	Gly	Lys	Arg
	序号	6	7	26	63	67	125	146
β-链	正常	Glu	Glu	Glu	His	Val	Glu	His
	Hb-S	Val	Glu	Glu	His	Val	Glu	His
	Hb-C	Lys	Glu	Glu	His	Val	Glu	His
	Hb-G San Jose	Glu	Gly	Glu	His	Val	Glu	His
	Hb-E	Glu	Glu	Lys	His	Val	Glu	His
	Hb-M Saskatoon	Glu	Glu	Glu	Tyr	Val	Glu	His
	Hb Zurich	Glu	Glu	Glu	Arg	Val	Glu	His
	Hb-M Milwaukee-I	Glu	Glu	Glu	His	Glu	Glu	His
	Hb-D β Punjab	Glu	Glu	Glu	His	Val	Gln	His

图 7－14　人类血红蛋白不同位置上的氨基酸的改变导致产生不同的贫血症。
(参考 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 8－25)

第四节　基因的精细结构和顺反测验

在经典遗传学中把基因看成是世代相传的遗传物质单位，它即是一个突变单位（从 *A* 突变为 *a*），又是一个重组单位，重组时整个的基因和另一个基因重组，而并不涉及到基因内部的重组；基因同时又是一个功能单位，它可以使机体显示出一定的表型。20世纪50年代中期本泽(Benzer)首先把"分子"基因的新概念引入了遗传学，此时双螺型的模型已经提出，Hexshey 等证实了噬菌体的遗传物质是 DNA，而不是蛋白质。血红蛋白单个氨基酸的改变导致其功能的改变也提示基因的突变仅仅部分位点的改变，而不是整个发生了变化。这些重大发现对基因精细结构的测定和顺反子概念的提出都具有一定的影响。更为直接的影响是①**拟等位基因**(pseudoallele)的提出。所谓的拟等位基因表型相似，但在位置上并不等位，②Benzer 于 1953 年发现 T－偶数噬菌体 rII(快速溶菌，rapid lysis)突变体和野生型 rII⁺在不同的 *E.coli* 宿主中生长情况和噬菌斑不同。突变型 rII 在 *E.coli* K 菌株中不能生长，仅能在 B 菌株中生长，噬菌斑大，边缘清楚；而野生型 *rII*⁺在两种菌株中都能生长，噬菌斑都是小而边缘模糊，和突变型明显不同（图 7－15）。（噬菌体的杂交和噬菌斑的形成见第 9 章）。看起来这一发现并不起眼，但本泽却利用这一现象建立了基因精细结构的测定方法和顺反子的概念。

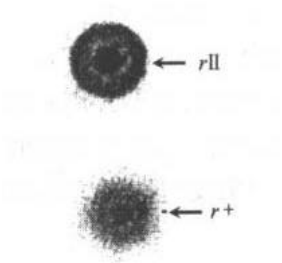


图 7－15　野生型 r⁺和突变型 rII 的噬菌斑。
（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 7－21）

一、基因的精细结构

1955 年 Benzer 发表了"细菌噬菌体遗传区的精细结构"的论文。他将相同的两个突变型混合感染 *E.coli* B 菌株，然后收集裂菌液，和 K 菌株混合后倒在固体平板上，结果在 K 菌株的菌苔(lawn)上观察噬菌斑的有无和多少。当两个突变型都是 *r*47 时，在 B 菌株上可以生长，但无法重组，裂解液中的噬菌体仍是突变型的，在 K 株菌中不能生长，所以在菌苔上没有噬菌斑(图 7－16a)。当将 *r*47 和 *r*109 同时混合感染 B 菌株后都能生长，由于突变的位点不同可以重组而获得野生型和双突变 *r*47，*r*104 的后代，只有野生型在 K 菌苔上可以生长，因此 K 菌苔上的噬菌斑的数目表示重组后野生型重组子的数目(图 7－16b)。b 组的结果只有少数噬菌斑，说明 *r*47 和 *r*104 两位点距离较近，重组的机会较少。C 组结果说明 *r*47 和 *r*106 的距离较远，d 组实验说明 *r*47 距 104 较前二者更远，所以噬菌斑最多。因此我们用以下公式可计算相邻不同位点的遗传距离：

$$Rf = \frac{2 \times r^{+}\text{重组噬菌体数}}{\text{总噬菌体数}} \times 100\%$$

重组的噬菌 r^{+} 数是 K 菌苔上的噬菌斑数。因为还有一半的重组后代是双重突变型，在 K 菌苔上不能生长，为了得到整个重组的后代数，故将 K 菌苔上的 r^{+} 噬菌斑乘

2。总的后代噬菌体数我们可以在将裂解液涂布 K 菌苔的同时将同样浓度的裂菌液涂布在 B 菌株菌苔上，由于突变型和野生型都能生长，所以噬菌斑数可以表示总的噬菌体后代数。那么以上公式就可以写成为

$$Rf = \frac{2 \times K\text{ 菌苔上的噬菌斑数}}{B\text{ 菌苔上的噬菌斑数}}$$

以前用来计算重组的方法是计算基因与基因之间的距离，而用本泽的方法可以用来测算一个基因内部不同位点的距离，它的精确度可达到十万分之一，即 0.001，也就是 0.001 个图距单位，故称为**基因的精细作图**。

二、顺反测验

本泽于 1857 年在 McElroy 等主编的 《遗传的化学基础》一书发表了遗传的基本单位的文章，提出了**顺反子**(cistron)的概念，他所做的实验也是利用 *r* 突变型和野生型的宿主差异的原理，但程序和精细作图实验正好相反（图 7－17)。他将两个不同的 *r* 噬菌体先混合感染 K 菌株，而不是 B 菌株。而且构成两个位点的不同排列(顺式和反式)，再比较**顺式**(*cis*)构型(a)、(c)和**反式**(*trans*)构型(b)、(d)的实验结果。*r*106 和 *r*51 无论是顺式排列还是反式排列在 K 菌株中都能生存，那么这两个突变型一定可以互补。这是由于 *r*106 和 *r*51 虽然相距较近，但都分居于 A、B 两个不同的顺反子中(图 7－18)，这样对于 A 顺反子来说不论顺式还是反式只有其中的一个等位基因内部发生了突变，而另一个等位基因是正常的，所以有功能。对 B 顺反子来说也是同样的道理。所以不论顺式还是反式都有功能(图 7－21a)。*r*47 和 *r*106 两个不同突变型混合感染 K 菌株时顺式有功能，是由于这两个位点同在 A 顺反子的一个等位基因内 (c)，而另一等位基因是正常的，所以有功能，当反式排列时，虽然两个位点相距较远，但仍在同一个顺反子内，两个突变位点分居于 A 顺反子的两个等位基因上，这两个基因都失去了

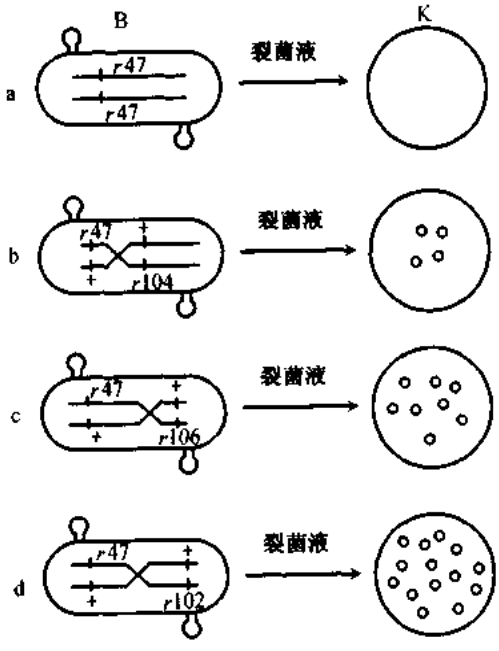


图 7－16 精细作图实验。

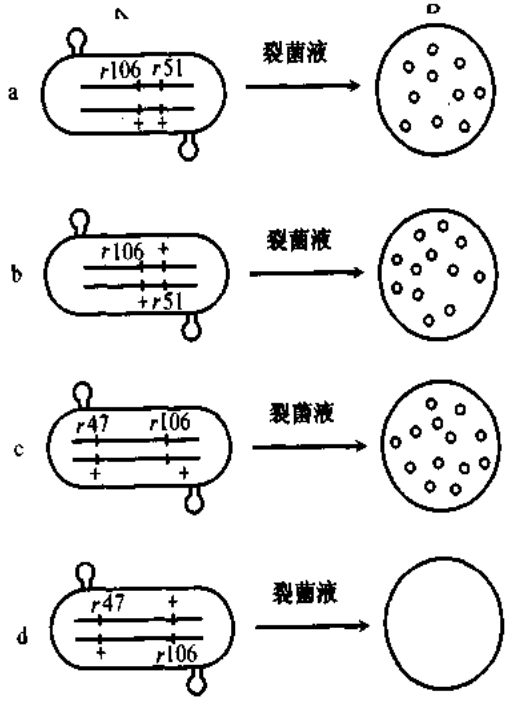


图 7－17 互补测验。

功能，综上所述，经过互补测验，表明，两个突变位点若可以互补，即顺式和反式排列都有功能，那么这两个突变位点不在同一个顺反子内；若两个突变位点不能互补，即顺式有功能，反式没有功能说明这两个突变位点同在一个顺反子内。一个顺子就是编码一个多肽的 DNA 序列，和基因的概念相同，但和经典的基因概念不同，顺反子仅是一个功能单位，其内部任何一个位点的突变都可能导致其功能的改变。其内部的不同位点之间也可以重组，所以它不是重组单位和突变单位。顺反子的提出使基因的概念向前发展了一大步，这也是遗传学从经典发展到分子遗传的重要标志之一。

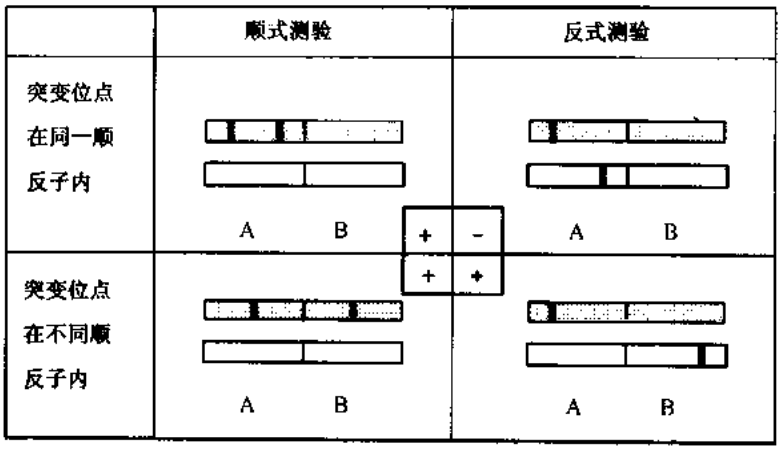


图 7－18　顺反测验结果的图解。当顺式有功能，而反式没有功能时，突变位点在同一顺反子内；当顺式有功能，而反式也有功能时，突变位点在不同顺反子内。（仿 Griffiths A. J. F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 12－39）

第八章

细菌和噬菌体的遗传重组

基本概念

转化(transformation)是指在生物之间通过细胞外 DNA 片段的途径来传递遗传物质。经遗传重组转化的 DNA 分子的一部分能和受体染色体的同源部分置换。转化可以用于确定基因顺序和基因间距离。

转染(transfection)是指用裸露的噬菌体整体 DNA 转化原核受体细胞的过程。用于真核细胞时,对任何裸露 DNA 的吸收都称为转染。在酵母和植物中转化和转染的概念是相似的。

接合(conjugation)是通过供体和受体菌细胞间的直接接触能单向转递遗传信息的过程。供体品系的细胞中有一个称为 F 因子的质粒。接合的结果是将 F 因子从供体转移到受体中。

F 因子可整合到细菌的染色体中,产生高频重组 Hfr 品系。Hfr 品系能和受体品系接合并转移供体的染色体。利用接合可以通过供体基因进入受体且发生重组的相对时间来进行细菌的基因作图。

转导(transduction)是以噬菌体为媒介将细菌的 DNA 从供体转移到受体中的过程。转导可以通过小段染色体来进行细菌精细结构作图。

用重组作图的遗传学原理可以用于不同的基因的突变位点之间的距离作图——基因间作图(intergenic mapping)及同一基因内部突变位点的作图——基因内作图(intragenic mapping)。

性导是通过 F′因子通过接合将某个特异的基因传递给受体的过程。

转导分为广泛性转导和特异性转导两种。广泛性转导是以 P1 或 P22 为媒介,通过"偷渡"和同源重组可将供体的任何基因传递给受体。两个基因间同时被 P1 转导的现象称为共转导。共转导可用于细菌的精细作图。特异性转导是以 λ 噬菌体为媒介而通过定点整合或重组将特异的基因(*gal* 或 *bio*)传递给受体的过程。

低频转导(LFT)是指通过缺陷型 λ(λdg)以较低的频度将 *gal* 基因传递给受体。

高频转导 (HFT) 是指含有 λ 和 λdg 的双重溶原菌为供体可将 *gal* 基因以很高的频率传递给受体。

第一节　基因和基因产物的符号

研究微生物的重组必定要涉及到基因及其产物、表型和基因型。1983 年大肠杆菌的基因位点已有 1027 个,因此必须要有一个统一稳定的符号才便于描述,1966 年 *Genetics* 刊登了 M. Demerec 等建议的大肠杆菌中基因符号的命名系统。1981 年 Sherman 又在此基础上提出了酿酒酵母基因符号的命名规则。现在国际各权威性生命学科杂志也制定了明确的要求,这些要求基本上和以上两大命名系统相一致。综合起来,规定如下:

大肠杆菌和其他细菌:基因座的命名统一用斜体,用三个缩写字母表示;表型第一个字母大写,用正体。基因型三个字母都小写,用斜体。肩上的符号用以表示野生型、突变型、抗性或敏感性,如 Gal^+ 表示表型为半乳糖野生型或原养型,Gal^- 或 Gal 表示表型为半乳糖突变型或缺陷型;gal^+ 表示基因型为半乳糖野生型,gal^- 或 *gal* 表示基因型为半乳糖突变型;Amp^r 表示表型为氨基苄青霉素抗性,Amp^s 表示表型为氨基苄

青霉素敏感性。*amp*^{r} 表示基因型为氨基苄青霉素抗性，*amp*^{s} 表示基因型为氨基苄青霉素敏感性。如一个基因有多个位点时，在基因座名称后用正体的大写字母来表示不同的位点，如 *lacZ*。对于特定的突变型常以它们被分离出来的前后顺序编号来表示，但编号要正写，如 *gal* K32。

酵母：三个字母表明基因功能，而后的数字表示不同基因座。酿酒酵母基因 *GAL4*，*CDC28*；蛋白质 GAL4，CDC28。非洲粟酒酵母基因 *gal4*，*cdc2*；蛋白质 Gal4，Cdc2。

线虫：用三个小写字母表示突变表型，如存在多个基因座，用连字符后接数字表示。例如，基因 *unc-86*，*ced-9*；蛋白质 UNC－86，CCED－9。

果蝇：突变表型的描述可以由 1～4 个字母表示。例如，基因 *white*（*w*），*tailless*（*tll*），*hedgehog*（*hh*）；蛋白质 White，Tailless，Hedgehog。

植物：虽然没有惯用法适用于所有植物，但大多数用 1～3 个小写字母表示。*Arabidopsis* 基因用果蝇的方法命名但需要用大写字母。例如，基因 *AGAMOUS*（*AG*）；蛋白质 AGAMOUS。

脊椎动物：一般以 1－4 个小写字母和数字表示其基因功能。例如，基因 *sey*，*myc*；蛋白质 Sey、Myc。

表 8－1 常用氨基酸的缩写符号

缩写		氨基酸		密码子
A	Ala	alamine	丙氨酸	GCU，GCC，GGG
C	Cys	cysteine	半胱氨酸	UGU，UGC
D	Asp	aspartic acid	天冬氨酸	UAG，UAC
E	Glu	glutamic acid	谷氨酸	GAA，GAG
F	Phe	phenylalanine	苯丙氨酸	UUU，UUC
G	Gly	glycine	甘氨酸	GGU，GGC，GGA，GGG
H	His	histidine	组氨酸	CAU，CAC
I	Ile	isoleucine	异亮氨酸	AUU，AUC，AUA
K	Lys	lysine	赖氨酸	AAA，AAG
L	Leu	leucine	亮氨酸	UUA，UUG，CUU，CUC，CUA，CUG
M	Met	methionine	甲硫氨酸	AUG
N	Asn	asparagine	天冬酰胺	AAU，AAC
P	Pro	proline	脯氨酸	CCU，CCC，CCA，CCG
Q	Gln	glutamine	谷氨酰胺	CAA，CAG
R	Arg	arginine	精氨酸	CGU，CGC，CGA，CGG，AGA，ACG
S	Ser	serine	丝氨酸	UCU，UCC，UCA，UCG，AGU，AGC
T	Thr	threonine	苏氨酸	ACU，ACC，ACA，ACG
V	Val	valine	缬氨酸	GUU，GUC，GUA，GUG
W	Trp	tryptophan	色氨酸	UGG
Y	Tyr	tyrosine	酪氨酸	UAU，UAC

人类：方法如脊椎动物但需大写。例如，基因 *MYC*、*ENO1*；蛋白质 MYC、ENO1。对于基因产物的命名以前没有统一的规定，现在基本上都统一用正体，或全部大写，或第一个字母大写，如 GAL 或 Gal。

表 8-2 细菌中用的基因符号

ara	阿拉伯糖不能利用	*pur*	嘌呤缺陷型(不能合成)
att	原噬菌体附着点	*pdx*	吡哆醇不能利用(不能合成)
azi	叠氮化钠抗性	*pyr*	嘧啶缺陷型(不能合成)
bio	生物素缺陷型(不能合成)	*rha*	鼠李糖不能利用
gal	半乳糖不能利用	*str*	链霉素抗性
lac	乳糖不能利用	*thi*	硫胺缺陷型(不能合成)
mal	麦芽糖不能利用	*ton*	噬菌体T1抗性
man	甘露糖不能利用	*tsx*	噬菌体T6抗性
mtl	甘露糖醇不能利用	*xyl*	木糖不能利用

第二节 细菌转化

遗传物质在细菌中有三种转移的形式，即**转化**(transformation)、**接合**(conjugation)和**转导**(transduction)。这三种转移形式的共同特点是：①单方向转移；②都产生部分二倍体(partial)；③基因只有整合到环状染色体上才能稳定地遗传。

转化是通过外源的 DNA 片段传递遗传信息的过程。此外源的 DNA 片段一般都是人工抽提的，传递时不需要载体，而是利用改变细胞膜的通透性，使外源 DNA 片段进入细胞，整合到基因组中，改变基因型。在细菌中，转化是裸露的非病毒 DNA 导入细胞的过程。**转染**(transfection)是转化的一种特殊形式，用于原核细胞时其外源 DNA 是特指离体的噬菌体 DNA，来感染感受态的细菌，并在其中得到表达。用于真核细胞时，对任何裸露 DNA 的吸收都称为转染。

细菌转化的现象是在 1928 年被 F.Griffith 首先发现的。1944 年 O.Avery 及其同事确定细菌转化因子是 DNA (见第 10 章)。在转化中游离 DNA 的片段被活的细菌所接受，而导致部分后代遗传性状发生了稳定的改变。转移过程通常用来进行制作细菌的基因图，这种作图方法是接合和转导所不能取代的。在转化作图的实验中，从一个供体品系的细胞中抽提和纯化 DNA，然后将它断裂成小的双链片段，再加入到不同基因型的**受体**(recipient)细菌的悬液中。被受体细胞所接受的**供体** (donor) DNA 片段就可以和受体染色体的同源部分进行重组，产生重组染色体，由于基因的重组导致了受体表型发生了改变，这个过程就称为转化。在转化中，任何品系都可以作为供体或受体，但供体和受体的基因型应是不同的。

大部分种类的细菌都可以通过转化而重组。但变异的关键在于它们吸收供体 DNA 的能力。有的细菌如枯草杆菌(*Bacillus subtilis*)在试管中能很容易被转化，用这种方法来作图。但野生型 *E.coli* 并不容易转化，这是由于细菌产生一种酶能迅速降解进入的外源 DNA。经过多年的努力，科学家们发现了一种方法可以增加细胞吸收外源 DNA 的效率。那就是用化学方法处理细胞，使其改变膜对 DNA 的通透性。这种细胞就称为**感受态细胞**(competent cells)，即细胞处于能摄入核酸分子时的生理状态。这种方法已经成为基因工程的常规技术，它对于我们利用体外 DNA 重组技术来了解真核和原核生物的基因功能特别重要。细菌的转化有两种类型：一种是**自然转化** (natural transformation)，在自然转化中细菌可以自由地吸收 DNA，通过它来进行遗传转化；另一种是**工程转化** (engineered transformation)，在这种转化中，细菌发生改变使得它们能摄入并转化外源 DNA。枯草杆菌中的转化属于自然转化，*E.coli* 转化属于工程转化。

在转化实验中，并非所有的受体细胞都被转化了，实际上只有很少比例的细胞吸收了外源 DNA，一旦供体的 DNA 被受体摄取，我们是可以检测到重组的转化产物和受体细胞的染色体。让我们来看一下枯草杆菌转化的例子(图 8－1)。

受体染色体上有一个突变基因 a，而供体的双链片段上有其野生型等位基因 a^+；外源 DNA 双链一条被降解，仅一条被摄入受体细胞(图 8－1/b)；这条单链线状 DNA 片段和受体的环状染色体的同源部分形成了三链结构(图 8－1/c)；供体的单链 DNA 和受体双链 DNA 通过双交换进行重组(图 8－1/d)；结果在双交换区域内，受体染色体发生重组，形成杂合双链 a/a^+，带有供体的 a^+ 基因，而被取代的单链上却带有受体的

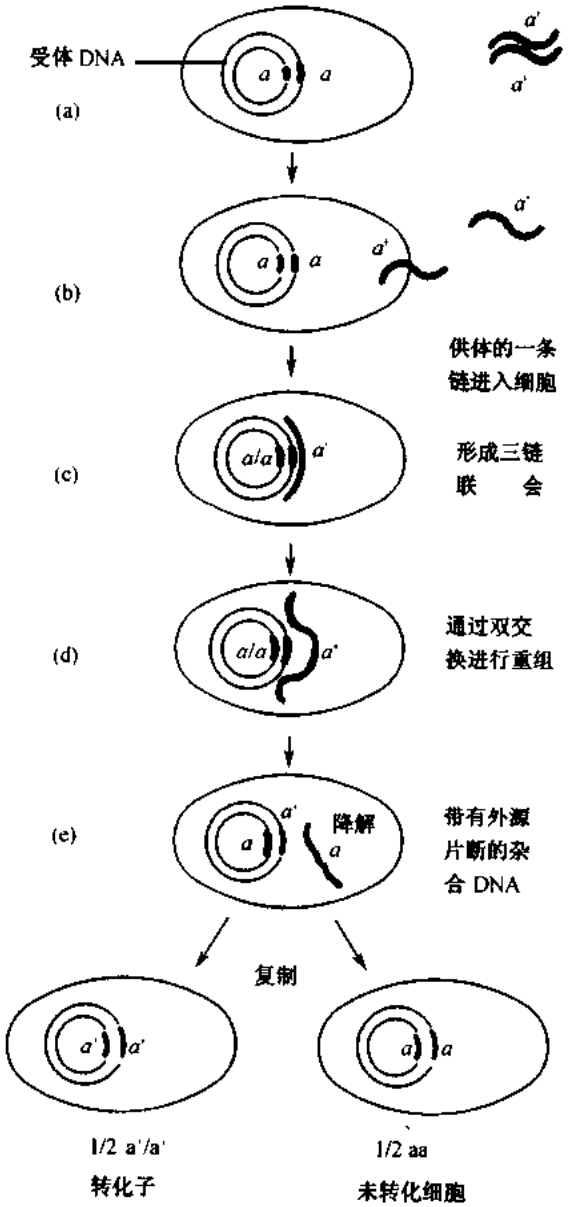


图 8－1　细菌转化的机制。(仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 7－3)

等位基因 a，后被降解(图 8－1e)；经复制后产生了两种不同的后代，一种是 a^+/a^+，另一种是 a/a，预计各占一半。对于高度感受态的受体细胞用过量的外源 DNA 片段进行转化，转化的频率约为 10^{-3}，即 10^3 个细胞中约有一个被转化。

通过转化可以测定基因的连锁、基因的排列顺序以及图距。原理如下：如果在供体的染色体上有两个离得很远的基因 a^+ 和 b^+，我们将会发现它们总是在不同的 DNA 片段上，这样如果有一个 a^+b^+ 供体和 ab 受体，那共转化（contransformation）即两个基因同时转化的概率是由两个基因单独转化的概率的乘积。若每个基因的转化产生率是 10^{-3} 的话，那么这两个基因同时被转化的频率应为 $10^{-3} \times 10^{-3} = 10^{-6}$。如果两个基因离得很近，有可能位于同一个 DNA 片段上，那么它们共转化的频率将接近于单个基因转化的频率。我们也可以这样说，若共转化的频率要比两个单个基因转化频率相乘积高的话，那么这两个基因一定是紧密连锁的。

基因的顺序也可以通过共转化的结果来分析确定，例如（图 8－2），p^+ 和 q^+ 常常共转化，而 q^+ 和 o^+ 也常常共转化，但基因 p^+ 和 o^+ 从未发生共转化，那么这三个基因的顺序一定是 p^-q^-o。

由于遗传学家可以控制转化片段的大小，因此两个基因共转化的频率可以和转化片段的平均大小相等。通过转化片段平均大小相关的共转化频率测定，就可以获得这些基因的连锁图谱。

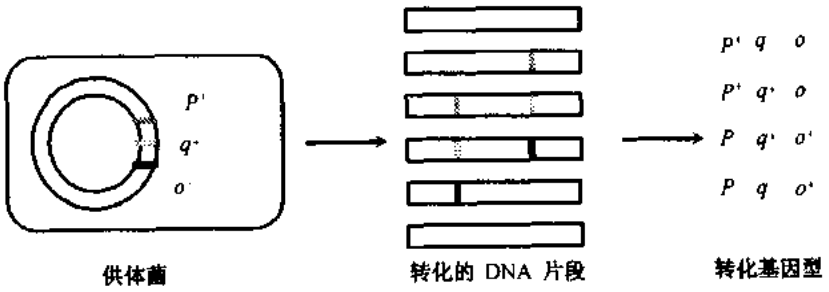


图 8－2 通过共转化确定基因顺序。（参考 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 7－4）

第三节 细菌的接合

一、*E. coli* 接合的发现

1946 年莱德伯格（J. Lederberg）和塔特姆（Tatum）发现在细菌之间转移遗传物质的另一种方法——**接合**（conjugation）。细菌接合是指通过细胞的直接接触，遗传信息从供体单向转移到受体的过程。

在 Lederberg 发现这种方法以前，人们已经对细菌的杂交进行了探索，但存在两大难题不易解决。① $a^+b^- \times a^-b^+ \rightarrow a^+b^+$ 杂交后得到的重组子（a^+b^+）的频率很低，和基因回复突变频率相近，难以区别。也就是说难以确定 a^+b^+ 是由 a^+b^- 或 a^-b^+ 亲本回复突变而产生，还是真正的重组子。② *E. coli* 杂交产生重组频率为 10^{-7}，也就是说要检查千万个菌落中才能发现一个重组菌落，工作量太大。Lederberg 采用了两个措施便巧妙地解决了这两个难题。一是使用**多重突变型**（multiple auxotroph）品系作为亲本进行杂交，由于两个或多个基因在同一亲体中同时发生回复突变的频率很小很小，这样可以十分信服地将重组子和回复突变区分开来。二是采用了以基本培养基来筛选原养

型的方法。由于制备多重突变型亲本工作量很大，正好在此不久前斯坦福大学的 Tatum 分离出了各种 *E.coli* 的生化突变型，于是 Lederberg 便邀请了这位曾和比德尔合作建立"一基因一酶"学说的微生物专家合作。巧的是 Tatum 就在 *E.coli* K12 品系中挑选了多重突变型，而正好 K12 中是有 F 因子，这是事前并不知道的，只不过 K12 是他在教学工作中多年使用的品系。如果当时要是采用了不含 F 因子的品系，那么实验难以成功，至少接合的发现要延迟数年。因此他们俩的这次合作在科学史上是一桩幸运的事，由于他们的特殊贡献，1958 年 Beadlr，Tatum 和 Lederberg 获得诺贝尔奖。他们选用两个 *E.coli* K12 多重突变品系：

A 品系：met^- bio^- thi^+ leu^+ thr^+（*bio*：生物素，*thi*：硫胺素 B1）

B 品系：met^+ bio^+ thi^- leu^- thr^-

将它们分别接种于基本培养基上都不能生长。若将它们混合后，经离心洗涤，除去完全培养基，再制成悬液以对照组相同的浓度接种于基本培养基，结果获得了野生型。其出现的频率为 10^{-7}。*E.coli* 许多生化突变型回复突变的频率是 10^{-7}，假如由亲本 A 两个位点同时回复突变成野生型的话，其概率应为 $10^{-7} \times 10^{-7} = 10^{-14}$；若 B 品系三个位点同时回复突变的话，其概率应为 $10^{-7} \times 10^{-7} \times 10^{-7} = 10^{-21}$，与重组子出现的频率 10^{-7} 相去甚远，故可以排除回复突变产生原养型的可能性（图 8－3）。

虽然回复突变的可能完全排除了，但人们仍提出以下质疑：①细菌的杂交实验获得的重组子可能是转化的结果。因为有可能一种亲本的细胞在处理过程中发生破裂，释放出转化因子，和另一亲本重组而产生部分的原养型。② 培养基中含有某些代谢产

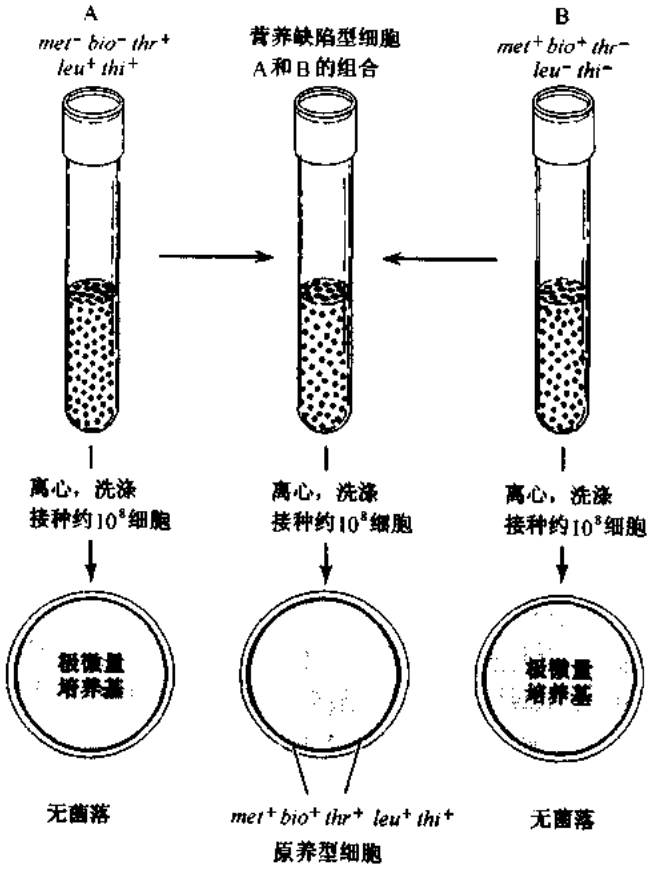


图 8－3　Lederberg 和 Tatum 的细菌杂交实验。

（引自 Russell，P.J.：Genetics，3rd ed.，1992，Fig.7－5）

物，混合后这些产物互相补充了对方的不足而得以在基本培养基上生长。这两个问题是尖锐而深入的。为澄清这个问题，莱德伯格和塔特姆将一种亲本的培养基经过细菌漏斗过滤，除去细胞，然后再加入另一亲本的培养物中，结果并未产生原养型。如果杂交实验中可以产生转化因子即 DNA 片段或某些可以互补的产物，那么它们是可以通过过滤器进入滤液中，进一步和受体作用，产生原养型，但结果正相反，说明也可以排除以上质疑，并提示 A 和 B 品系细菌的直接接触似乎是出现原养型的必要条件。1950 年戴维斯(B.D.Davis)设计了 U 形管实验，进一步支持了上述结论(图 8-4)。U 形管的底部中内有一砂芯横隔，上有微孔，可以通过 0.1μm 以下的颗粒，细胞无法通过。戴维斯将 A、B 品系分别放入 U 形管的两侧，一侧塞上棉塞，另一端接上注射器。当两种细菌培养物都增生到饱和状态时，用注射器轻轻地把培养液从一臂经过滤器吸到另一臂，再轻轻压过去，让培养液充分混匀，但两种品系的细胞无法接触，然后两侧的 A、B 菌株离心洗涤后再涂布到基本培养基上，都不出现菌落，即不产生重组子。若培养液中有转化因子 DNA，或代谢产物的话是可以在两个品系之间充分交流的，但并未引起重组，因此完全可以排除转化和代谢产物互补的可能性，而清楚地表明，两个亲本的细胞不互相接触是不能进行杂交的，这为研究接合的本质奠定了基础。由于历史的局限，从一开始莱德伯格和塔特姆就根据真核的重组来解释他们的发现，认为细菌接合是一个彼此交换遗传物质的过程，即细菌**同宗配合**(homothallic)。导致 1951 年莱德伯格根据自己的实验，将 *E.coli* 的连锁图绘成三叉戟状，使 *E.coli* 连锁图的绘制陷入困境。任何一个伟大的科学家在他整个的一生中，难免会出现一些失误，但他们那种创新的精神，求实的态度，永远值得后人的学习和赞美。瑕不掩瑜用在他们身上是十分合适的。

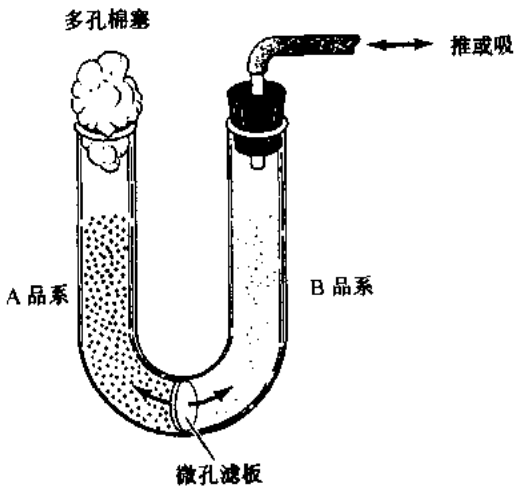


图 8-4　戴维斯的 U 型管实验。
(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 7-6)

二、单向转移与 F 因子

1952 年海斯(W. Hayes)的意外发现，动摇了“细菌同宗配合”的观点，使细胞接合的研究又转向正确的方向。

海斯在重复莱德伯格和塔特姆的细菌杂交实验之前，分别用高剂量的链霉素来处

理 A 品系和 B 品系，结果 A 品系受处理后，杂交结果不受影响，而 B 品系受处理后却完全阻止了重组的发生。这对海斯来说意味着两个品系在杂交的过程中作用是不相同的，也就是说可能不是交互的过程，而是一个单向的过程。因为链霉素可以阻止细胞分裂，继而杀死细胞。但它还允许交配持续一个短时间。A 品系在交配后被杀死并不影响杂交结果，表明它是一种供体，像雄性动物一样，交配后被杀死，不会影响后代。而 B 品系像是一种受体，和雌性动物一样相似，受孕后被杀死的话，就无法产生后代。海斯的结论认为细菌的接合是**异宗配合**（heterothallic）过程，两个亲本在杂交中所起的作用不同，有供体和受体之分，相当于雄性和雌性。

以上结论得出不久，海斯将放在冰箱中有一年之久的一个菌种 A（str^s）♂ 和 B（str^s）♀ 进行杂交，A 品系就假设为雄性♂，B 品系假设为雌性♀，结果和预期的不同，不能产生重组子，海斯对这一偶然的意外情况并没有掉以轻心，而是继续实验和分析。他将此菌进行诱变，使其由 str^s 变成 str^r，以便试着和 A（str^s）杂交，结果是可育的，产生了 A（str^r），再进一步和 B（str^s）品系杂交，结果是可育的，说明原来是“雄”性的 A 品系（图 8－5）。放在冰箱中一年，变成了雌性，即从供体变成了受体。以上实验完成后不久，莱德伯格和他的妻子 E．Lederberg 以及 L．L．Cavalli 也发现了这一现象，他们对此一致的解释是这样的（图 8－6），所谓的供体或雄性是细菌的细胞内有一种致育因子（fertility factor，F），以“F^+”表示，而雌性品系是缺乏这种 F 因子，以“F^-”表示。原先的雄性品系是“F^+”放置冰箱一年后丢失了 F 因子，变成了 F^-，作为 F^- 的 A 品系是不能和另一 F^- 的 B（str^s）杂交，所以不育。经诱变后其基因型由（str^s）变成了（str^r），但仍为 F^-，故可与 F^+ 的 A（str^s）杂交，经 *str* 选择培养基筛选出重组子仍是 F^+，和 F^-（str^s）杂交仍然可育，这样便发现了 F 因子。

F 因子是一种质粒（plasmid），由共价环状闭合 DNA 双链构成，全长 94.5kb，主要

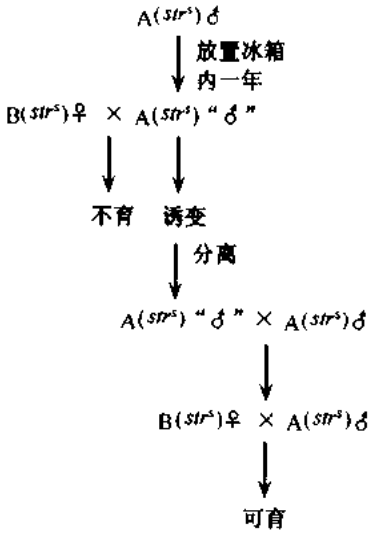


图 8－5　海斯发现的意外情况和实验。

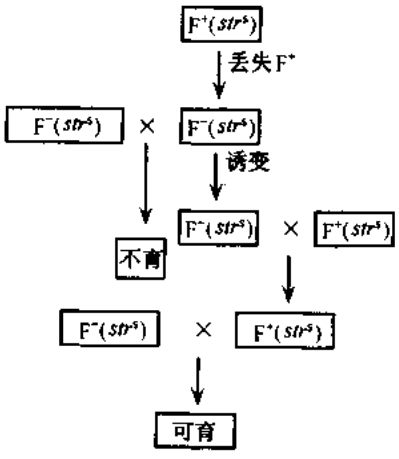


图 8－6　意外发现和实验结果的解释。

分为三个区域(图 8-7):①重组区;②自主复制区;③转移区。重组区中有 4 个插入顺序 (insertion sequence, IS),通过和宿主染色体上的 IS 同源重组或通过转座(transposition),F 因子可以整合到宿主不同位点上。由于宿主染色体上的 IS 方向不同,所以整合的方向也随之不同。自主复制区有转移的起点和 2 个复制起点,复制起点 ori T 是在染色体转移时进行滚环复制时的复制起点(图 8-7),ori V 是在营养时期,即游离在细胞质中独立复制时的复制起点。转移区主要有合成性伞毛即 F 纤毛(F pill)的操纵子。F 纤毛是由性伞毛蛋白构成,呈管状,又叫**接合管**。通过接合管可将供体和受体细胞相联。1957 年拍到了第一张接合照片。当时对 F 因子知道得并不透彻,也不知道重组体是怎样产生的,但从各种实验的结果推论出以下结论:①有 F 因子的细菌称为 F^+,细菌增殖时可把 F 因子传递给后代;②没有 F 因子的细菌称为 F^-,F^+细菌经吖啶橙处理而丢失,成为 F^-。F 因子一经丢失,细胞中便不再出现;③F^+可以和 F^-杂交,而不能和 F^+杂交;④$F^+ \times F^-$的杂交后代皆为 F^+,而且可以 10^{-7} 频率获得重组体后代。

F 因子介导的接合也和滚环复制有关。其**转移区**(transfer region)长 33kb。此是接合所需要的。这一段含有 40 个基因和 DNA 转移有关(整个基因组约有 60 个基因),其中

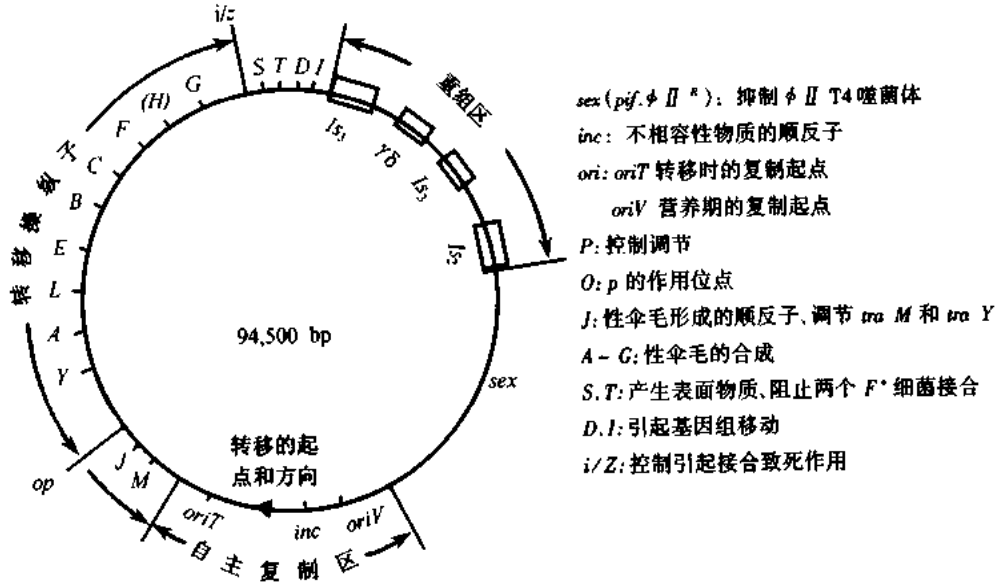


图 8-7 F 因子的结构。
(参考 G.Zubay:Genetics,p.485)

主要的基因及功能见表 8-3。这些基因叫 *tra*,它们主要的部分作为一个 32kb 的转录单位(*traY*-1)协同表达。*traM* 和 *traJ* 分开表达(图 8-8)。*traJ* 是调节 *traM* 和 *traY*-1 两者的打开,在另一条链上的 *finP* 是一个调节子,产生一个小分子反义 RNA,关闭 *traJ*。它的激活需要另一些基因 *finO* 的表达。转录单位中的某些基因直接和 DNA 的转移有关,但大多和细胞的特点有关。

细菌表面的**性伞毛**(pili)是由 F 因子编码的。*traA* 编码了伞毛蛋白(pilin)的单个亚基,由此聚合为伞毛。至少有 12 个 *tra* 基因是修饰和聚合伞毛蛋白的,为伞毛所必需。F 性伞毛为毛发状结构,长 2~3μm,由细胞表面伸出,一个典型的 F^+细胞有 2~3 根性伞毛。这些伞毛蛋白亚基聚合成中空的管子,直径 8 nm,中间有 2 nm 的轴孔。并不是遗传物质转移的孔道。供体细胞不会接触其他 F^+细胞,那是因为 *traS* 和 *traT* 编码了

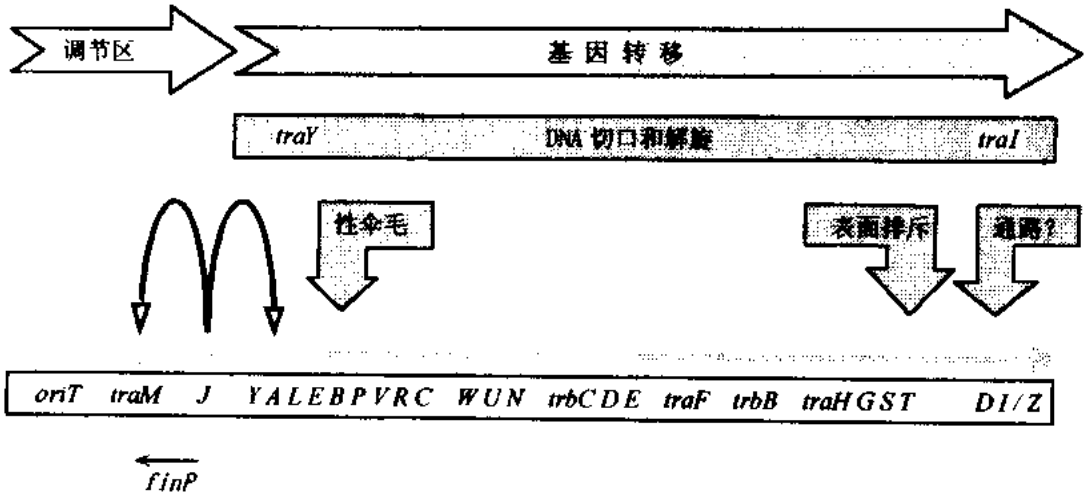


图 8－8　F 因子 *tra* 区和接合有关的基因（仿 B.Lewin: Genes, Ⅵ, 1997, Fig.14－20）
traJ 调节 *traM*、*traY*－*I* 打开。*finP* 是 *traJ* 的反义 RNA，由 *finO* 激活。*traA* 编码伞毛蛋白亚基，性散毛聚合蛋白。*TraS*－*T* 编码表面排斥蛋白。TraY/I 切开 *oriT*。

表 8－3　F 因子与转移功能有关的基因及功能

基因	基因产物的功能
traM	衣壳蛋白，可能涉及对受体F⁻细胞的识别
traJ	是*traM*和*tra*操纵子转录的正调控因子
finP	是traJ的反向转录调节物
traY	内切酶，在*oriT*附近产生缺口
traA	编码性伞毛蛋白
traL、*E*、*K*、*B*、*V*、*C*、*W*	性伞毛的装配和修饰
trbC、*traU*、*F*、*H*、*G*	性伞毛的装配和交配接合物的稳定
traN	同上
traS、*T*	表面排斥阻止和F′接合
traD	是F′的内膜蛋白，接合时涉及DNA转运，可能是孔道部分
traI	解旋酶性质的复制蛋白，可切开*oriT*的*nic*位点

表面排斥蛋白，这种蛋白使细胞有效地限制了 F′和 F′细胞接合。F 性伞毛还可以提供 RNA 噬菌体和某些单链 DNA 噬菌体，如 M52、R17、f1 等吸附位点。因此 F′细菌对这些噬菌体是敏感的，而 F⁻细菌是抗性的。在供体细胞和受体细胞开始接触时是易断裂的，但另外一些 *tra* 基因起到稳定联接的作用，使接合的细胞紧密联接在一起。性伞毛对起始配对是必要的，但一旦起始以后即使性伞毛收缩和消失，接合的细胞已彼此紧密接触了。DNA 转移时必须有一个通道通过。但性伞毛并不存在这种通道。因此这种通道仍未确认出来。但 *traD* 可能提供了通道或通道的一部分。

第四节 高频重组品系

一、高频重组品系的发现

就在 F 因子发现之前，Luca Cavalli-Sforza（1951 年）偶然发现，用氮芥处理 A 品系后从存活的细菌中分离出一种新的品系，将这种新品系和经典的 B 品系进行杂交所得到的重组体比一般的杂交要高出一千倍。这个品系就称为**高频重组**（High frequence recombination，Hfr）品系。1954 年海斯也分离出了 Hfr 品系，称为 Hfr H，Cavalli 分离出的 Hfr 品系就称为 Hfr C。

高频重组品系和经典的 F^+ 菌株的特点有相同的地方，也有不同之处，归纳起来可分成以下几点。相同处：①都能和 F^- 进行杂交产生重组后代；②杂交时都要通过接合管和受体菌相联接；③用高剂量链霉素处理后都不影响杂交，说明它们都是作为一种供体。不同处：①产生重组子的频率不同。$Hfr \times F^-$ 的重组子产生频率为 10^{-4}，而 $F^+ \times F^-$ 产生的重组子仅为 10^{-7}。②$F^+ \times F^-$ 的后代常为 F^+，而 $Hfr \times F^-$ 的后代皆为 F^-；③F^+ 经吖啶橙处理后会变成 F^-，而 Hfr 经吖啶橙处理后仍为 Hfr，可以和 F^- 杂交。

二、Hfr 和 F 因子的关系

F 因子既可以游离于细胞质中，也可以整合在宿主的染色体上，这种质粒就称为**附加体**（episome）或游离基因。当处于游离状态时，细菌为 F^+，当整合到宿主染色体上

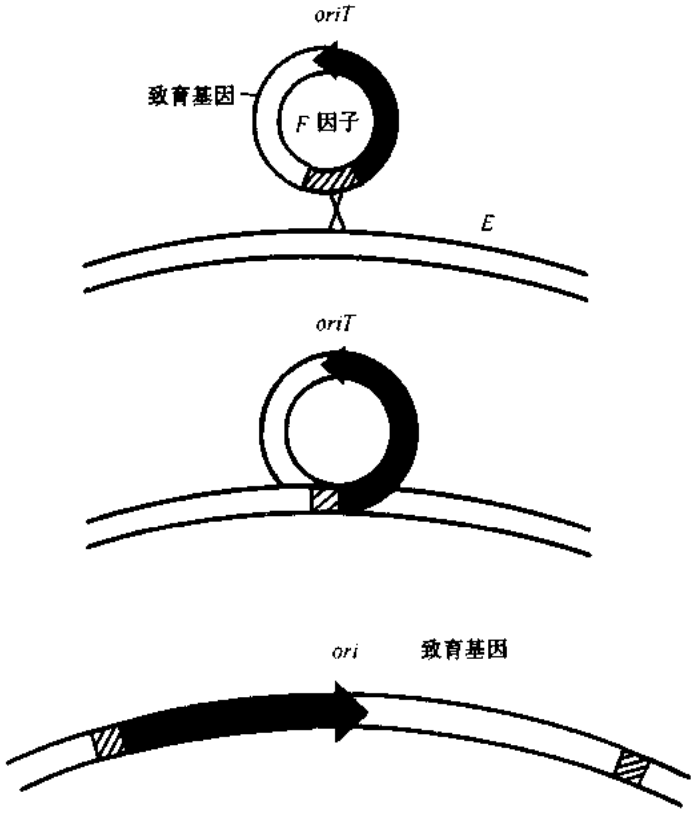


图 8－9 F 因子的整合。

时即为 Hfr 品系。所以经吖啶橙处理不会丢失 F 因子。Hfr 和 F^- 细胞接触时 *Ori T* 活化，进行滚环复制。开始时，缺刻蛋白识别并结合在 *ori T* 区，切开单链，以另一条链为模板，在 3′－OH 上从头合成。5′ 端沿箭头方向延伸，将宿主的 DNA 移到受体中。由于整合后致育基因位于 DNA 环的末端（图 8－9），一般尚未进入受体细胞前，接合管就已经断裂，使转移中断，故 Hfr 杂交的后代不能获得致育基因，故不能产生性伞毛而呈 F^- 的性状。$F^+ \times F^-$ 一般只能将 F 因子转入受体，使受体变成 F^+（图 8－10）。只有在先整合到宿主的染色体上形成 Hfr，才能在杂交时将宿主的基因转移到受体中，经重组产生重组子。F 因子整合的概率是 10^{-3}，$Hfr \times F^-$ 产生重组子的概率是 10^{-4}，所以从 $F^+ \times F^-$ 产生重组子的频率是 10^{-7}。

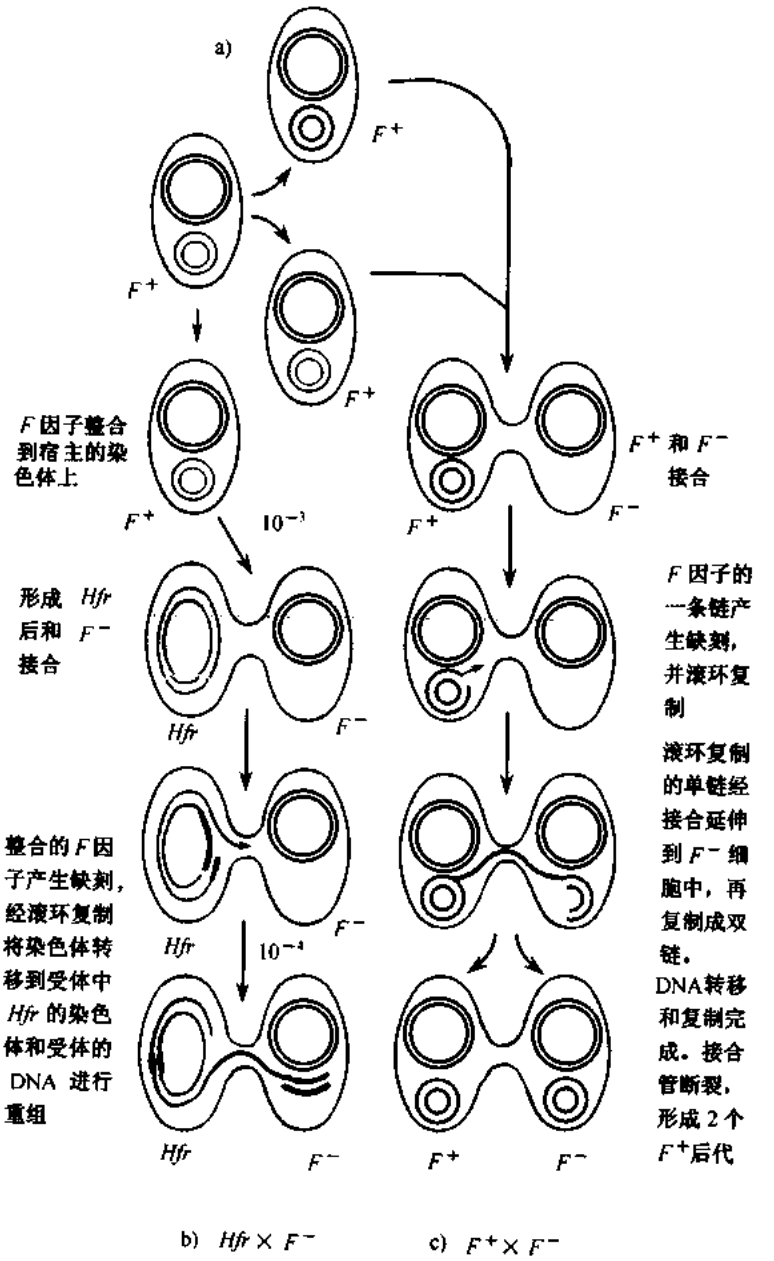


图 8－10　*F* 因子和 *Hfr* 的关系。

三、性导

1959年E. H. Adelberg和Burus在重复 *Hfr* × F^- 实验时，发现有一些品系回复成了 F^+ 状态，失去了高频供体的能力，即不能将供体的DNA片段转移到受体细胞中。但F因子已带有基因，被称为F′因子。F. Jacob和Adelberg发现这种品系以很高的频率将 lac^+ 传递给受体，使受体的表型变成 F^+-lac^+。但后代并不稳定，约有 10^{-3} 子细胞又变成 F- lac^-，所以重组后代的基因型为 F-lac^+/ lac^-（图8-9）。在已知 *F′* 是由于整合在宿主染色体上的F因子，有时会环出，偶然发生错位，将整合位点附近的基因一道环出，形成了带有宿主片段DNA的F因子，这就是F′因子。当和 F^- 接合时，随着F因子的转移也将它所携带的基因导入受体，形成部分二倍体 F′ - lac^+/ lac^-。这种部分二倍体称为**局部二倍体**(merodoploidy)。F′ - lac^+/ lac^- 不稳定，由于F′因子的丢失(10^{-3})而变成 F′ ~ lac^- 细胞。

F′和F因子不同：①F′ 品系转变成Hfr品系的频率要高于 F^+ 品系；②F′ 变成Hfr时F′因子整合到相同位点上，而 F^+ 变成Hfr时可整合到不同位点；③F′ × F^- 高频传递特定的基因，形成部分二倍体，而 F^+ × F^- 要么产生 F^+ 但不转移任何基因，要么以 10^{-7} 将宿主的基因按顺序转入受体，但受体仍为 F^-。所转移的基因是不同的。

以F′因子为媒介将特定的基因转递到受体中的现象就称为**性导**(sexduction)。

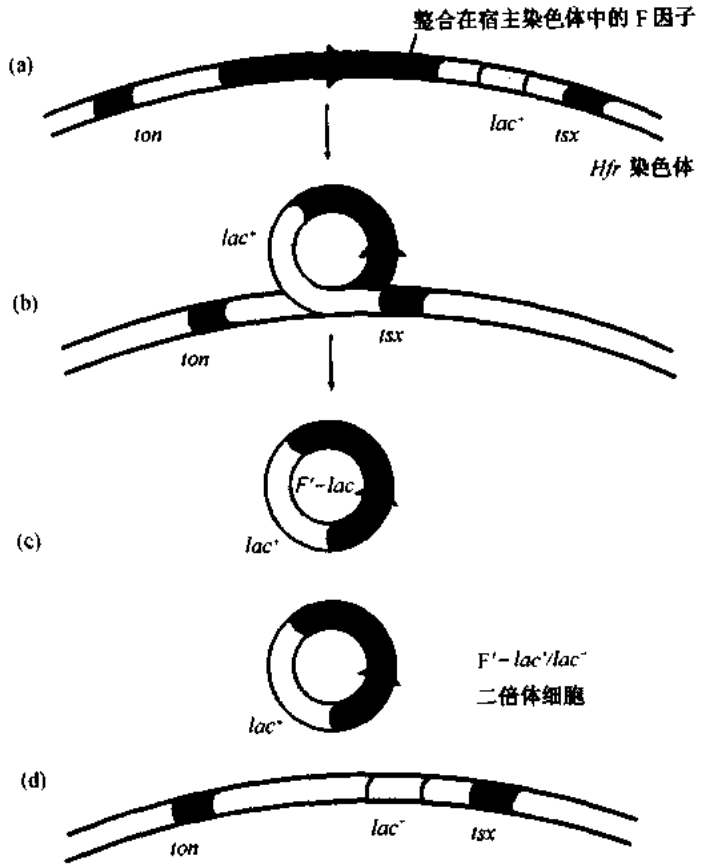


图8-11 F′因子的起源及部分二倍体的形成。（仿 Griffiths, A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10-15）

Jacob 和 Adelberg 发现了 F′ – *lac* 后，其他的因子又先后陆续被发现，其中之一是抗药性质粒，也称为 R 因子（resistance factor）。当时日本已使用抗菌素药物十几年之久，发现有些致病菌具有对抗菌素能产生抗药性，而且还可以把这种抗药性转移给其他敏感性菌株。T. Watanabe 和 T. Ukasawa（1960 年）经研究证实了抗药性传递是由于一种新的附加体所致，他们称之为**抗性转移因子**（resistance-transfer factor，RTF），即 R 因子。结构如图 8 – 12 所示，和 F 因子的结构相似，不同的是其中的转座因子（见第 23 章）中带各种抗药性基因，故可产生抗药性。

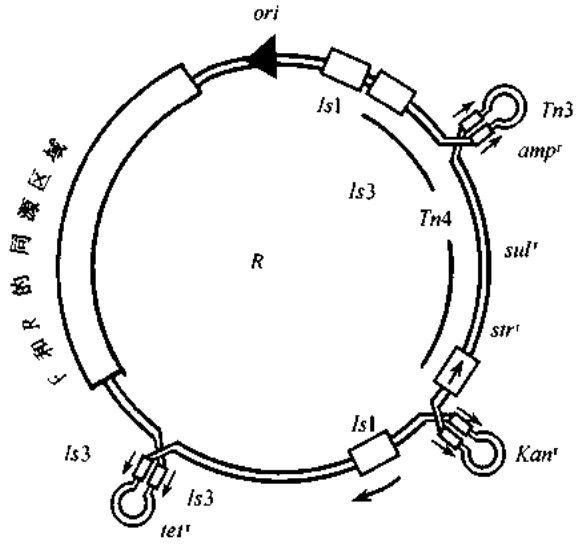


图 8 – 12　R 因子的结构。

四、重组频率的计算

我们可以用不同的方法来计算 *E. coli* 的各基因的重组频率，绘制出 *E. coli* 的染色体图。下面将介绍 3 种思路不同的计算方法：

1. 传递等级绘图

1956 年，Wollman，Jacob 和 Hayes 发表了 *E. coli* K12 的接合和遗传重组。他们用的两个亲本分别为：

Hfr H：　*thr*$^+$　*leu*$^+$　*azi*^{s}　*lac*$^+$　*gal*$^+$　(*λ*)$^-$　*str*^{s}　*mal*$^+$　*mtl*$^+$

F$^-$：　*thr*$^-$　*leu*$^-$　*azi*^{r}　*lac*$^-$　*gal*$^-$　(*λ*)$^+$　*str*^{r}　*mal*$^-$　*mtl*$^-$

他们采用了两组**选择标记**（selected marker），一组是 *thr*$^+$ *leu*$^+$ *str*^{r}，也就是说重组子中一定含有 *thr*$^+$ *leu*$^+$ *str*^{r}，即双交换的两个交换点一个必须在 *thr*$^+$ *leu*$^+$ 的左侧，而另一个必须在 *str*^{r} 的左侧。第二组是 *gal*$^+$ 和 *str*^{r}，那么双交换的两个交换点一个必须在 *gal*$^+$ 的左侧，另一个必须在 *gal*$^+$ 和 *str*^{r} 之间（图 8 – 13）。实验是这样进行的，将混合后的细菌悬液涂布在选择培养基上，选择培养基是在基本培养基中加了 Str 和其他物质，而不加 Thr、Leu 或 Gal 相应酶类。能在这种选择培养基上生长的菌株一定是 *thr*$^+$ *leu*$^+$ *str*^{r} 或 *gal*$^+$ *str*^{r}，而其他标记可正可负，这些皆**属非选择性标记**（nonselected marker）。将 *thr*$^+$ leu$^+$ str^r 或 *gal*$^+$ *str*^{r} 的重组子再经选择培养基，如在基本培养基中加入各种相关物质，

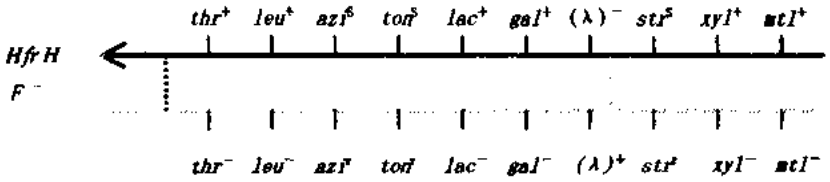


图 8-13 Hfr×F⁻杂交传递等级实验中双交换发生的位点。

表 8-4 *Hfr* × *F*⁻杂交中的传递等级

选择标记	非选择标记								
	thr⁺/*leu*⁺	*azi*⁶	*ton*ˢ	*lac*⁺	*gal*⁺	(λ)	*mal*⁺	*xyl*⁺	*mtl*⁺
thr⁺ *leu*⁺ *str*ˢ	–	92	73	49	31	15	0	0	0
gal⁺ *str*ˢ	75	75	74	74	–	84	0	0	0

参考 Wollman, Jacob 和 Hayes, 1956

惟独不加 Lac 相应酶类，能够在这种培养基上生长的重组子是 *thr*⁺ *lac*⁺ *str*ˢ 或 *gal*⁺ *lac*⁺ *str*ˢ，这种选择培养基上长出的菌落数与前一种选择培养基上的菌落数的百分比就是各种非选择性标记的百分率(表 8-4)，如 *lac* 的非选择性标记的百分率为

$$\frac{thr^+\ leu^+\ str^s\ \ lac^+}{thr^+\ leu^+\ str^s} \times 100\% = 49\%$$

Wollman 等的实验结果(表 8-4)表明各非选择性标记显示了不同的等级，离选择标记越近的基因(非选择标记)出现越高，反之越低，*myl*⁺(D-木酮糖利用) *mtl*⁺(甘露醇糖)座位在双交换范围以外，不可能在重组子中出现，故重组频率皆为零。

既然和选择标记的距离不同显示了一定的等级，那么反过来我们通过非选择标记重组频率的测定也能推断出遗传距离并进行作图。

2. 中断杂交作图

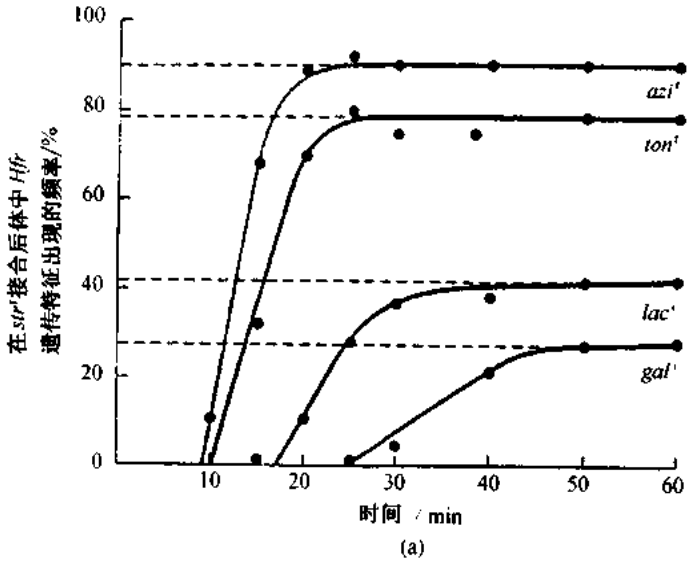
1961 年 Jacob 和 Wollman 又利用中断杂交的方法建立了另一种绘制 *E.coli* 物理图谱的方法。他们将 *Hfr* × *F*⁻杂交，在混合后不同时间进行搅拌，以使接合管断裂，中止染色体向受体细胞转移。越是排在前面的基因进入受体的机会越高，反之排在离起点较后的基因进入受体的机会越低，因此用不同时间来中断杂交，就能推断出基因的顺序，他们所用的亲本是：

供体： *Hfr H str*ˢ *thr*⁺ *leu*⁺ *azi*ʳ *ton*ˢ *lac*⁺ *gal*⁺

受体： *F*⁻ *str*ˢ *thr leu azi*ˢ *ton*ˢ *lac*⁻ *gal*⁻

交配后在不同时间将样品搅拌 2 分钟，然后稀释防止重新接合，再把它们涂布在含有链霉素的选择培养基上，选出 *thr*⁺ *leu*⁺ *str*ˢ，再按上述实验方法测定不同时间中断时各非选择标记的频率结果如表 8-5(图 8-12) 所示，*thr*⁺和 *leu*⁺是选择标记，首先进入受体细胞，形成**部分二倍体**(partiolly diploid)，即细胞具有全套的内源基因和部分的外源基因，经双交换形成**重组子**。*azi*ʳ 约在杂交后 9 分钟进入受体并重组，随着时间的增加，其相对重组率(*azi*ʳ / *thr*⁺ *leu*⁺ ×100%)迅速升高，这是由于时间延长，重组的机会增大了，20 分钟左右达到 90%，趋于平衡，不再上升。这一频率极值反映了 *azi*ʳ

和 *thr*$^+$ *leu*$^+$同时重组入受体染色体的频率，因此不可能达到 100%，而且和选择标记相距越近，同时重组的机会也越高。*ton*^{s} 约在 10 分钟进入受体细胞并重组，20 分钟后相对重组率接近 80%；*lac*$^+$约在 17 分钟，*gal*$^+$约在 20 分钟进入受体并重组。以上特点归纳起来有 3 点：①不同的非选择标记进入受体且重组的时间不同，但都是稳定的；②各基因的重组频率随着时间的增加而增加，直至极值；③按时间顺序，先重组和后重组的基因，重组率的极值在逐步减少；此是由于上述等级传递所致。



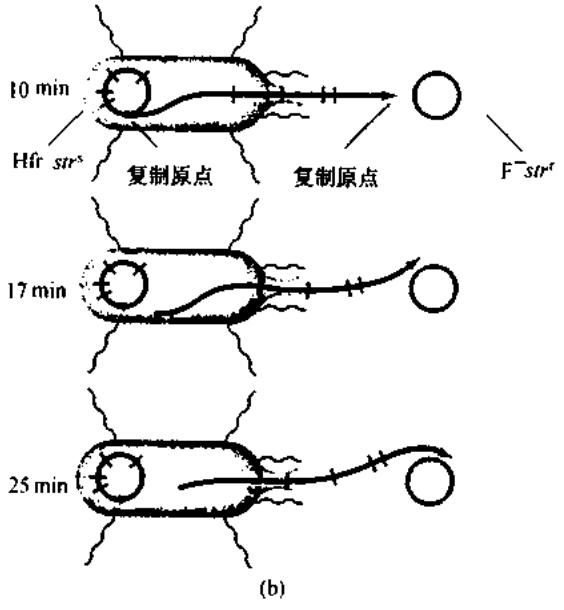


图 8－14　*E.coli* 的中断杂交实验。*str*^{r} 品系的 F$^-$细胞与 *str*^{s} 品系的 *Hfr* 杂交。F$^-$细胞有很多的突变(以遗传标记 *azi*、*ton*、*lac* 和 *gal* 来表示)，而 *Hfr* 品系的相应标记都是野生型的。当两种细胞混合后在不同的时间中断两个细胞之间的接合，然后再分别接种于不同的选择培养基上，进行检测。(a)在杂交后不同时间各标记的重组频率图；(b) 在不同时间标记转移的图示。（转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10－6）

表 8-5 *Hfr* × *F*⁻中断杂交中不同标记时间非选择标记标记的重组率

取样时间/min	各非选择标记标记的重组频率/%				
	thr⁺ *leu*⁺	*azi*ʳ	*ton*ʳ	*lac*⁺	*gal*⁺
5	0	0	0	0	0
10	100	12	3	0	0
15	100	70	31	0	0
20	100	88	71	12	20
40	100	90	75	38	20

参考 Jacob 和 Wollman。

根据以上实验，意味着特定基因进入受体的时间和该基因在染色体上的排列顺序成正比例关系。因此利用中断杂交实验也可以进行 *E.coli* 的染色体作图（图 8-15）。中断杂交作图的结果和传递等级绘图的结果是一致的。

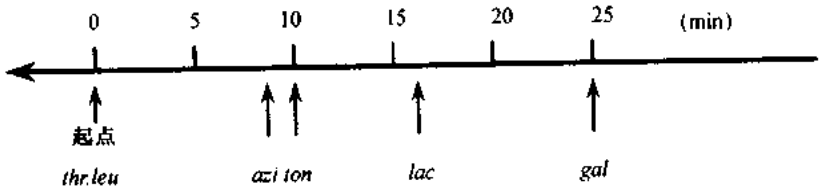


图 8-15 利用中断杂交进行 *E.coli* 染色体作图。

3. 重组作图——*E.coli* 染色体连锁图

接合重组的过程有 2 个要素，一是基因要能在供体、受体分开前进入受体细胞；二是要和选择标记一道通过双交换整合到受体的基因组中才能稳定地表达和遗传。在原核细胞中重组并非发生在两个完整的基因组之间，而是在**部分二倍体**（partial diploid）中，部分二倍体或**部分合子**（merozygote）也称**半合子**，那就是细胞中具有完整的基因组（称**内基因子**）和来自供体的部分基因组（**外基因子**，exogenote）（图 8-16）。若外基因子和内基因子之间发生一次单交换，那么产物是线状的（图 8-16 b），不仅不能复制，且易降解，所以是不稳定的，只有发生双交换时（图 8-16 c），才能产生稳定的环状重组体，和一个片段，同样这个小片段也将被降解掉。

绘制染色体图就必须较精确地计算基因之间的距离，中断杂交虽然能以转移的时间来确定基因的顺序和距离，但较为粗放，在 2 分钟之内就难以精确测定。同时测定重组就要以同等机会进入受体的转移 DNA 来进行测定，但由于交配的自发中断，这就产生了一个自然的转移梯度，越是排在后面的基因，转移到受体的机会越小，例如供体的 3 个标记排列为 *met*⁺ *arg*⁺ *leu*⁺，可以预期到受体的片段分配如图 8-17。

不难看出靠近起点的基因在接合中有较多的机会转移。因此利用转移的 DNA 梯度也可确定基因排列的顺序。

在了解了以上的一些原理后，就有可能以相对的重组频率为依据来构成一张有意义的遗传图。于是 Jacob 和 Wallman 进行了 Hfr H *str*ˢ × F⁻ *pro*⁻ *lac*⁻ *pur*⁻ *str*ʳ 杂交实验，以 *pur*⁺ *str*ʳ 为选择标记，以 *pro*⁻ *lac*⁻ 为非选择标记，实际上就是含有 *pro*，*lac*，*pur*，str 4 个标记的等长片段来分析重组，这些片段进入受体的机会是相同的。当将 *pur*⁺ 和

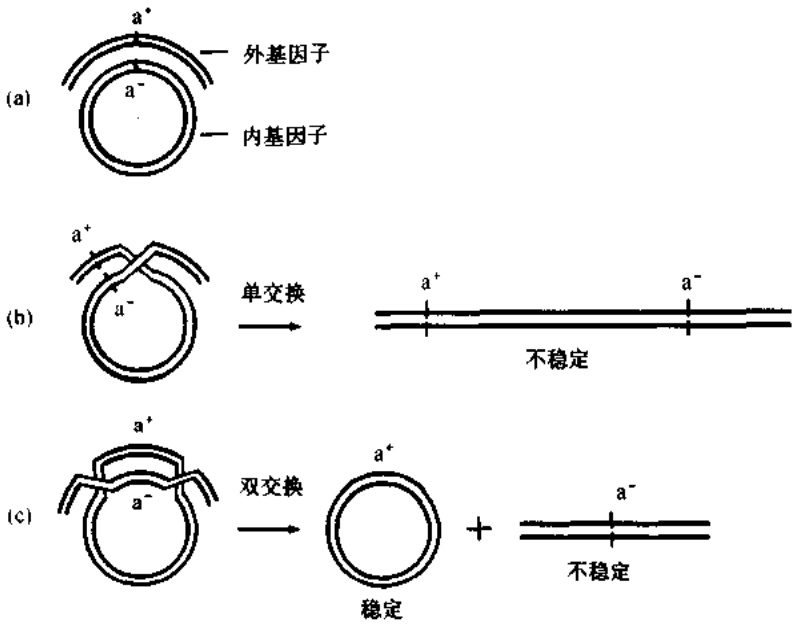


图 8－16　在部分二倍体中外基因子和内基因子之间的重组。（仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10－12）

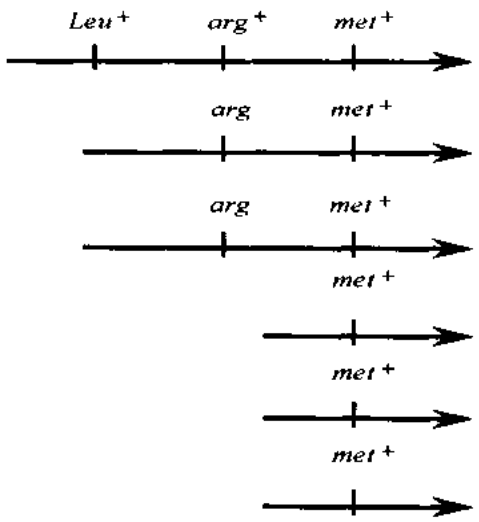


图 8－17　根据自然转移的梯度所预期的受体片段的分配。（引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993）

距很远的基因不易处于一个转移片段上。

str^r 为选择标记时，重组有四种类型。和真核生物的重组道理相同，两个基因相距越远的话，它们之间发生交换的机会越大，反之越小。在四种类型中 *str*^{r} *pur*$^+$ *lac*$^-$ *pro*$^-$ 发生的机会应最小，其相对的重组频率为

$$\frac{pur^+\ lac^-\ pro^-}{pur^+} \times 100\% = 22\%$$

分母 *pur*$^+$ 代表 *str*^{s} *pur*$^+$ 的所有重组子，即图 8－18 的四种类型。此重组值可以表示 *pur* – *lac* 之间的图距，转移时间约为 1 分钟。也就是说，中断杂交作图中的 1 分钟距离，相当于 22 m.u.。用这种方法对短距离的测定是可靠的，但不能用于远距离的测定，因相

五、*E.coli* 环状染色体

F 因子整合到 *E.coli* 的染色体上有 20 多个不同的整合位点，而且整合的方向也可不同，这是由宿主染色体上的 IS 的位置和方向所决定的。这样染色体转移的起到和方向也都不同，因而形成了不同的品系（图 8－19）。

不同品系基因的顺序反过来也说明 *E.coli* 的染色体是环状的，否则就不可能产生

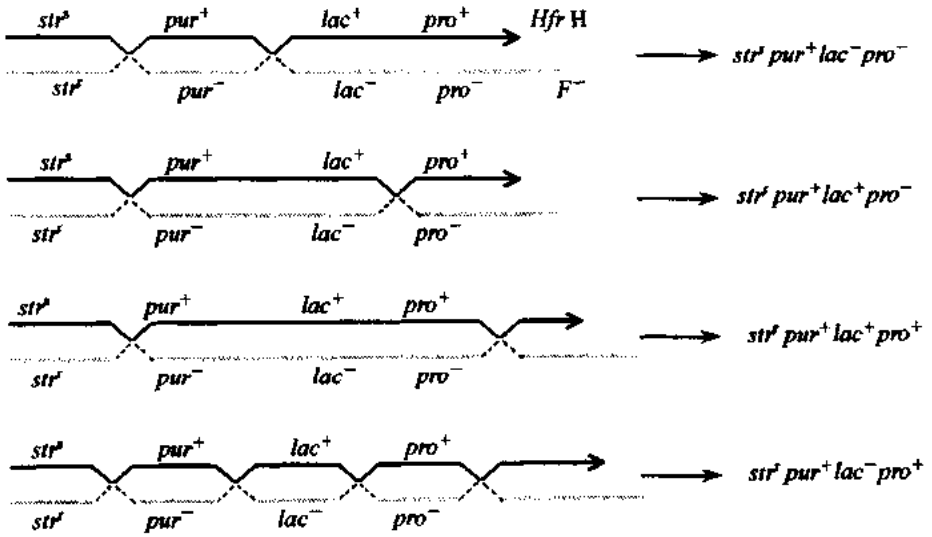


图 8－18　*Hfr*H *str*^{s} × F$^-$ *pro*$^-$ *lac*$^-$ *pur*$^-$，以 *str*^{r} *pur*$^+$ 选择标记的重组类型。
（仿 Griffiths A.J.F. *et al.*：An Introduction to Genetic Analysis，5th ed.，1993，Fig.10－13）

这种情况。通过接合人们将获得的资料已绘制出 *E. coli* 的环状染色体(图 8－20)。

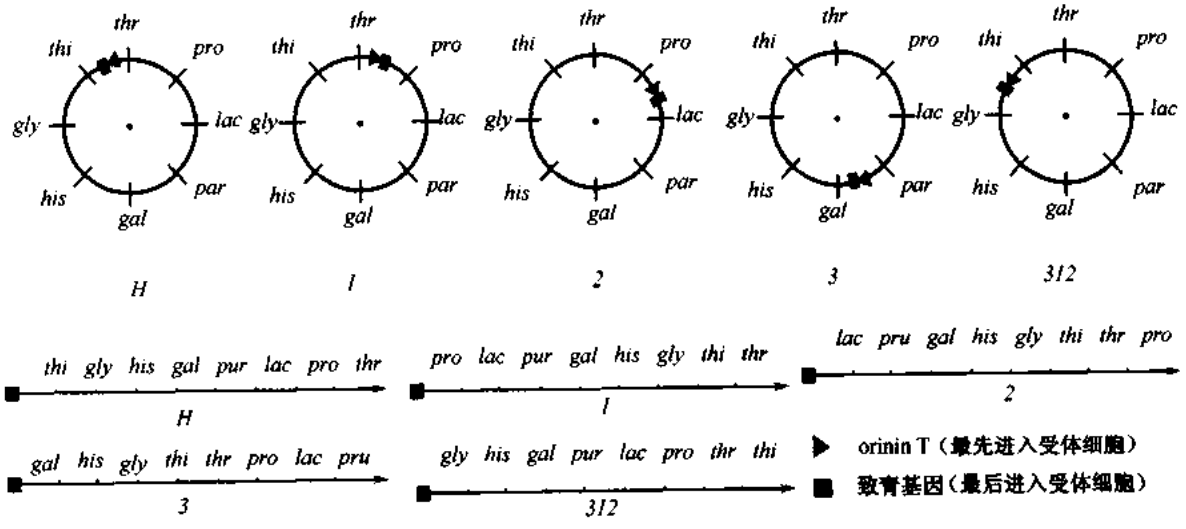


图 8－19　*E. coli* 的环状染色体。上：F 因子处于染色体上的不同位点；下：高效重组时基因转移的先后顺序不同。
（仿 Griffiths A.J.F. *et al.*：An Introduction to Genetic Analysis，50 ed.，1993，Fig.10－8）

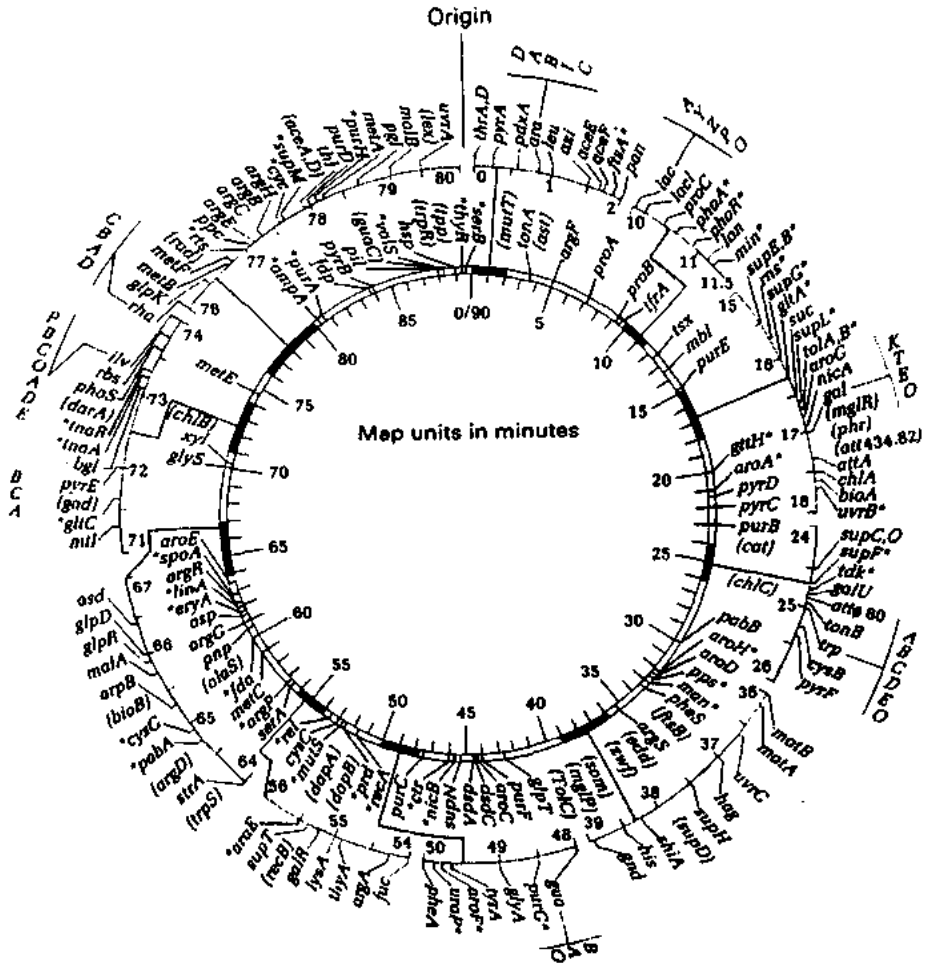


图 8－20　根据接合实验所作出的 *E. coli* 的遗传图谱。单位是分钟，起点处为零分钟。
（转引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 7－30）

第五节　噬菌体染色体的连锁和交换

一、噬菌体的结构和形态

噬菌体有三种不同形态：20 面体无尾的，如 PM－2；20 面体有尾的，如 SPO1；丝状的如 M13。有尾噬菌体一般有头、颈、核心、鞘、尾板、尾丝几部分组成（图 8－21）。噬菌体的遗传物质有的是 DNA，有的是 RNA；有的是双链，有的是单链；有的是线状，有的是环状（表 8－6）。

噬菌体的生活史分成两种途径——**裂解途径**（lytic pathway）和**溶原途径**（lysogenic pathway）。仅能裂解生长的噬菌体叫**烈性噬菌体**（virulent）。溶原反应只存在双链 DNA 噬菌体中，这类噬菌体称为**温和性噬菌体**（temperate），它们感染细胞质，并不复制。染色体整合到宿主的染色体中，此时的噬菌体称为**原噬菌体**（prophage）。带有原噬菌体的细菌称为**溶原性细菌**（lysogeny bacteria），因它可导致敏感性细菌裂解，故称“溶原”。

表 8-6 几种病毒染色体的特点

病毒	宿主	核酸结构	染色体类型	染色体长度/μm	GC含量/%
T-偶数噬菌体	*E. coli*	双链DNA	线环、环状排列末端有RS	60	35
T7	*E. coli*	双链DNA	线状;单一顺序	12	48
λ	*E. coli*	双链DNA	线状;单股粘性末端	16	49
p22	沙门氏菌	双链DNA	线状;单一顺序	14	48
$\phi X174$	*E. coli*	单链DNA	环状	1.8	A28 G24 T33 C18
Qβ	*E. coli*	单链DNA	线状	1.4	A22 G24 U29 C25
(呼肠病毒)	哺乳动物	双链DNA	几个片段	8.3	A38 G17 U28 C17
SV40	人类	双链DNA	超螺旋环	1.7	41
鼠白血病病毒	鼠	单链DNA	几个片段		A25 G25 U23 C27
烟草花叶病毒	烟草	单链DNA	线状		A30 G25 U26 C19

引自 Peter J. Russell: Genetics, 3rd ed., 1972, Table 10.1

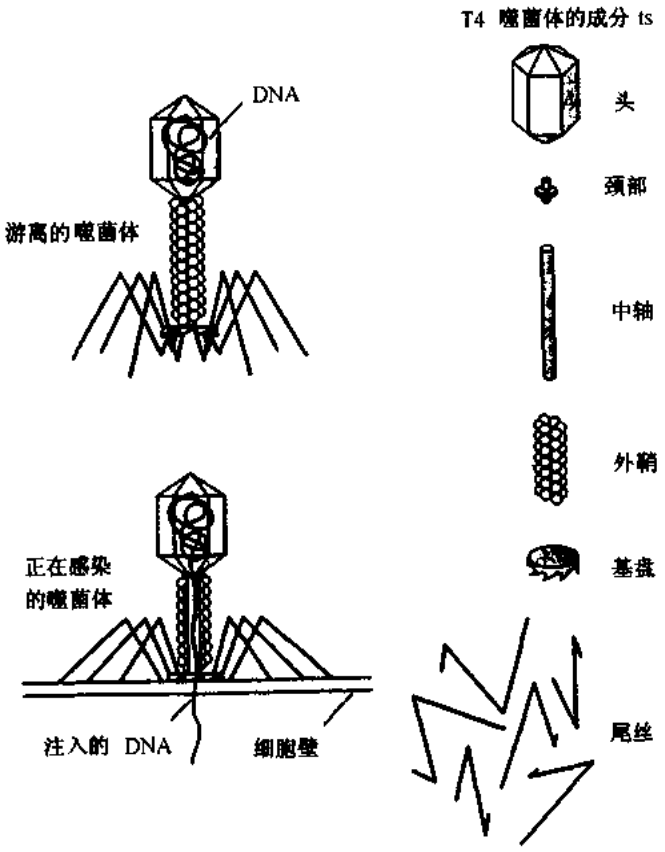


图 8-21 游离状态的 T4 噬菌体和感染 *E. coli* 时的过程。
(转引自 Griffiths A. J. F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10-17)

二、λ 噬菌体

1951 年 J. Lederberg 的妻子 Esther Lederberg 证明了 J. Lederberg 和 Tatum 用来杂交的 K12 中有原噬菌体,并命名为 λ,经 10 年的研究搞清了溶原化的实质。

λ噬菌体的基因组长达 50 kb，共 61 个基因，其中 38 个较为重要。其生活史如图 8－22 所示，可分为裂解周期和溶原周期。细菌处于溶原化状态时，细胞质中有一些 λ *CI* 基因的产物 CI 蛋白（见第 18 章），这是一种阻遏蛋白，可以阻止 λ 左、右两个早期起动子的转录，使之不能产生一些复制及细胞裂解的蛋白。λ 的 DNA 随着宿主的染色体复制而复制。但在 UV 诱导下 Rec 蛋白可降解 CI 蛋白（见第 17 章），诱导 90% 的细胞裂解。有时 λ 也可自发地（10^{-5}）从宿主的染色体上游离出来，进行复制，最终导致宿主细胞的裂解，此称为**治愈**（curing）。游离在细胞质中的 λ 可以进行滚环复制，产生多个拷贝，并合成头部和尾部蛋白，包装成完整的 λ 噬菌体，使细胞裂解，释放出 λ 噬菌体再感染新的细胞。（图 8－22）。因为 λ 噬菌体的 DNA 也有整合在染色体上和游离于细胞质中两种状态，所以也称做**附加体**。但和 F 因子不同，λ 噬菌体有细胞外形式，而 F 因子无细胞外形式。

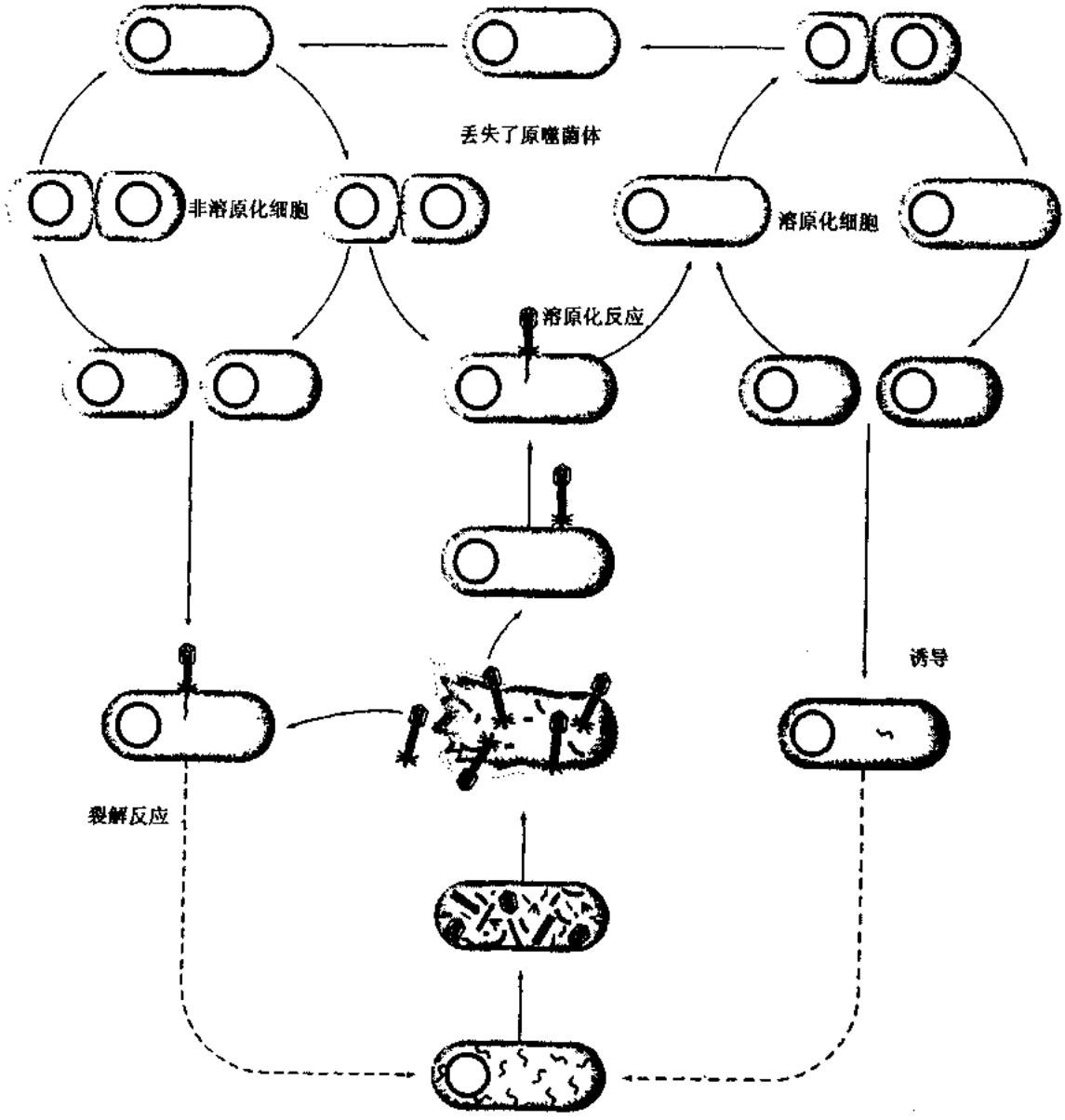


图 8－22　温和性噬菌体细胞周期的交替。

（转引自 Griffiths A. J. F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10－26）

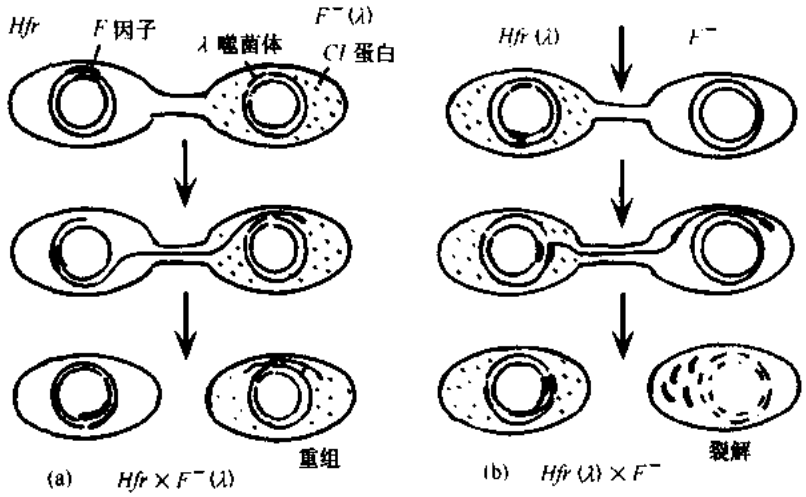


图 8-23 合子诱导的机理。

在 *E.coli* K12 中是有原噬菌体的存在。Jacob 和 Wollman(1956 年)发现了**合子诱导**(zygotic induction)现象，并利用合子诱导确定了几个 *E.coli* 染色体上原噬菌体的整合位点。他们发现 *Hfr*(λ) × F^-所得到的重组子频率要比 *Hfr* × F^-(λ)或 *Hfr*(λ) × F^-(λ)低得多。这是由于在 *Hfr*(λ) × F^-的杂交中，原噬菌体进入无阻遏物的受体细胞质中，进行大量复制使受体细胞裂解(图 8-22b)，因此不易得到重组子，此现象就称为合子诱导。现在我们再回过头来查阅一下传递等级作图，中断杂交实验以及重组作图都是采用 *Hfr* × F^-(λ)就是不致产生合子诱导的缘故。

三、噬菌体的基因重组

1936 年 F. M. Burnet 发表了噬菌体能产生突变体，其噬菌斑的外形和野生型的有明显区别，可惜的未能引起重视，以致噬菌体遗传学延迟了 10 年才得以建立。

1946 年第 11 届冷泉港学术讨论会上，在宣布一基因一酶学说的胜利，及 Lederberg、Tatum 细菌杂交实验报告的同时，Hershey 和 Luria 宣布发现了噬菌体的 *r*、*h* 突变，Delbrück 和 Hershey 发表了他们各自发现的噬菌体重组，这四项重大的发现分别在 1958 年和 1969 年获得了诺贝尔奖。后两项的发现有力地推动了噬菌体遗传学的发展。

噬菌体的基因重组和细菌不同，而和真核的重组十分相似。杂交是用标记不同的噬菌体之间进行。然后计算重组噬菌体占总的子代噬菌体的比例来确定重组值。一般可以选用 2～4 个基因差异的噬菌体来混合感染细菌。首先把不同类型的噬菌体混合起来和细菌一起涂布在固体培养基上，细菌的浓度要达到可以长成**菌苔**（lawn）的水平，噬菌体的浓度要很稀。每个噬菌体感染一个细菌，经过裂解周期，宿主细胞破裂后，释放出的子噬菌体又去感染周围的细菌，结果在菌苔上形成一个圆形清亮的斑，称为**噬菌斑**（phague），而一个噬菌斑来自最初涂布平板时的一个噬菌体。噬菌斑的形态必须选择容易区别的，以表示噬菌体的相应表型。单个的噬菌体只能在电镜下才可观察其形态，突变引起其形态变化没有电镜是无法鉴别的，但突变影响到生活周期，会产生不同的噬菌斑，因此通过噬菌斑的观察我们很容易观察基因型的变化与重组。

Hershey 等用 T2 噬菌体的两个不同表型特征：噬菌斑的形态和宿主范围来进行

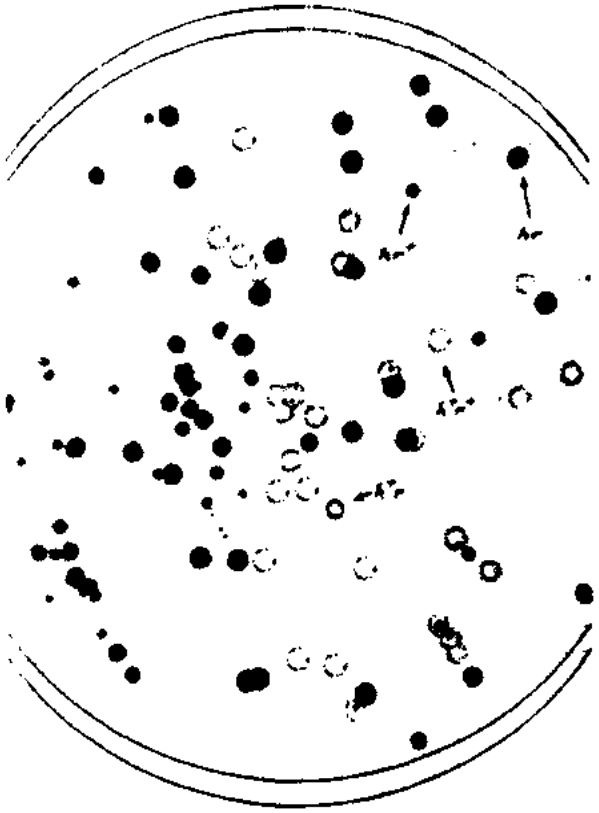


图 8－24　hr^+ 和 h^+r 混合感染 *E.coli* 和 B/2 时在混合菌苔上出现的四种噬菌斑。
（转引自 Russell、P.J.:Genetics,3rd ed.，1992,Fig.7－20）

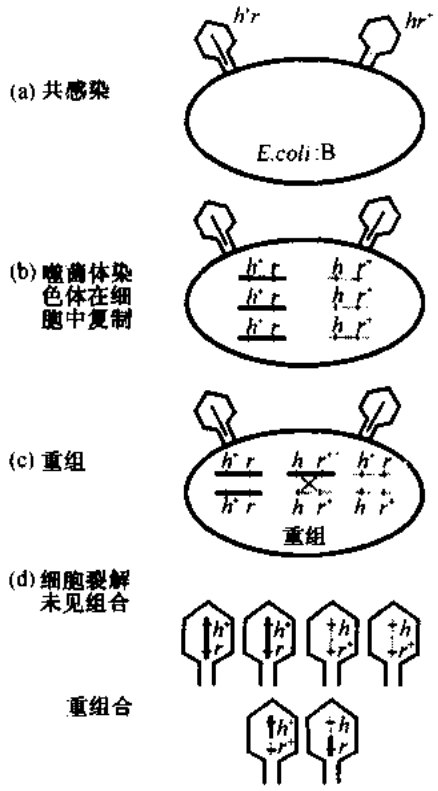


图 8－25　$h^+r \times h\ r^+$ 的杂交。

杂交。一个噬菌体的基因型是 h^+r，另一个噬菌体的基因型是 $h\ r^+$。h^+表示宿主范围（hostrange），是野生型，能在 *E.coli* B 菌株上生长，r 表示快速溶菌（rapid lysis），产生的噬菌斑大，边缘清楚。h 噬菌体能在 *E.coli* B 和 B/2 品系上生长，r^+产生小而边缘模糊的噬菌斑，若将 h 噬菌体倒在含有 *E.coli* B 及 B/2 的菌苔上，能产生透明的噬菌斑，而 h^+ 因只能裂解 *E.coli* B，所以在 B 和 B/2 的混合菌苔上产生的噬菌斑是半透明的。

杂交时 hr^+和 h^+r 混合感染 *E.coli* B 和 B/2，在 B 和 B/2 混合菌苔上出现了四种噬菌斑（表 8－7）（图 8－24），表明 $h\ r^+$ 和 h^+r 之间有一部分染色体在 B 菌株的细胞中进行了重组，释放出的子噬菌体有一部分的基因型为 h^+r^+和 $h\ r$。我们利用下面的公式就可以计算出和两个位点的重组值：

$$\text{重组值} = \frac{h^+r^+ + h\ r}{\text{总噬菌斑数}} \times 100\%$$

此重组值也表示两个连锁基因之间的遗传距离。

表 8-7 $hr^+ \times h^+r$ 混合感染

E. coli B和B2的结果		
表型		推导的基因型
透明	小	hr^+
半透明	大	h^+r^+
半透明	小	h^+r^+
透明	大	hr

四、T2的环形遗传图

Hershey 根据 $h\ r^+ \times h^+\ r$ 的杂交结果推测 T2 的染色体是线状的，其实并非如此。在完成了上述 T2 杂交实验后，Hershey 又分离出了许多 T2 噬菌体，详细地研究了 T2 噬菌体的遗传重组。杂交结果如表 8-8 所示。根据实验的结果 r_1，r_7 和 r_{13} 的连锁图（图 8-26）有四种不同的排列方式，有的 r_1, r_7, r_{13} 居于 h 的同侧，有的分居两侧，分居两侧的情况也各有不同。究竟那一种是正确的呢？这必须要进一步掌握 r_1，

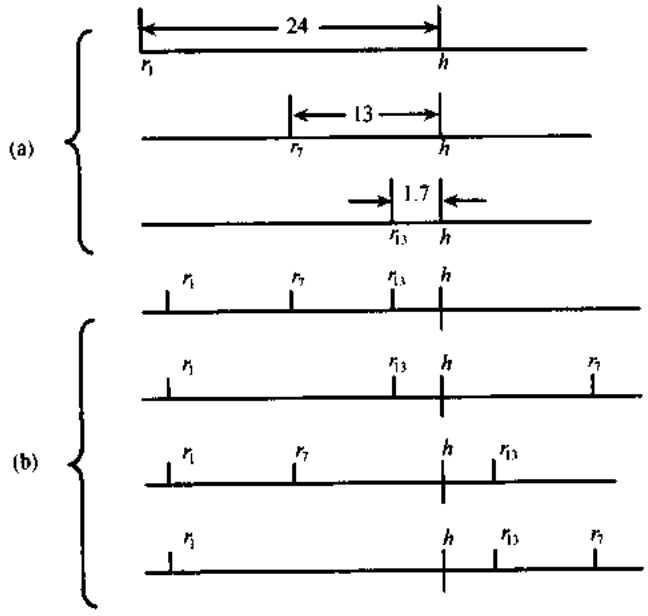


图 8-26 rh 的图距和 r_1、r_7、r_{13} 的不同排列。（仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10-24）

r_7 和 r_{13} 相互间的距离才能确定。进一步通过 $r_{13}r_7^+ \times r_{13}^+ r_7$ 的杂交，确定其排列为 $r_{13}^-\ h^-\ r_7$，再通过一系列的杂交结果发现第一种和第四种排列都是可能的，Hershey 当时无法判明这两种可能中哪一种是正确的，10 多年后 G. Streisinger 和 Edgar(1964 年)研究了噬菌体多个标记基因的杂交重组，才判断出 T-偶数噬菌体的遗传图是环形的，所以上述的两种可能排列都是正确的。也只有在染色体是环形的情况下，才能合理理解 Hershey 的实验结果。T4 噬菌体的环形遗传图如图 8-27 所示。

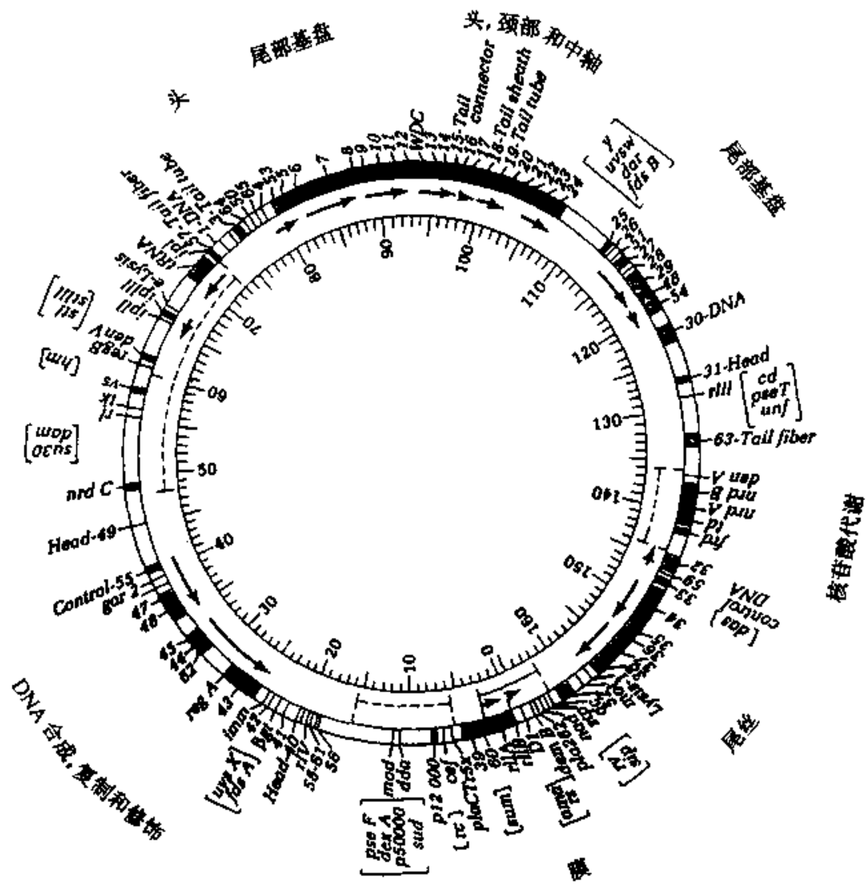


图 8-27　噬菌体 T4 的遗传图谱。注意图的外周所示的是相关基因簇的功能。单位是 kb。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 7-31）(1kb = 1000 碱基对)

表 8-8　$T_2hr_+ \times T_2h_+r$ 杂交的重组率

杂交	各种基因型的百分比/%				重组率/%
	h^+r^+	hr^+	h^+r	hr	
$hr^+ \times h^+r_1$	12	42	34	12	24
$hr^+ \times h^+r_7$	5.9	56	32	6.4	13
$hr^+ \times h^+r_{13}$	0.74	59	39	0.94	1.7

第六节　转　　导

有的噬菌体能将一个细菌的遗传物质带到另一个细菌中这一过程就称为**转导**(transduction)。转导有 2 种：**普遍性转导**（generalized transduction）和**特异性转导**(specialized transduction)。普遍性转导的噬菌体能携带供体染色体上的任何部分，而特异性转导只能携带细菌染色体的特异部分。

一、转导的发现

1952 年 J. Lederberg 和他的学生 N. Zinder 在鼠伤寒沙门氏细菌（*Salmonella ty-*

phimurium) 中寻找遗传重组，所用的方法和 *E.coli* 的杂交实验相似。他们用 *phe*⁻ *trp*⁻ *tyr*⁻ × *met*⁻ *his*⁻ 结果获得了 10^{-5} 的原养型重组子(*phe*⁺ *trp*⁺ *tyr*⁺ *met*⁺ *his*⁺)。但将两个亲本放在 U 形管的两臂中，结果也能同样获得重组子，此和 *E.coli* 杂交是不同的，于是他们推论以上的杂交并不需要细胞的直接接触，而可能是以一种滤过性因子(F_A)为媒介来传递遗传物质。根据 FA 和沙门氏细菌中的温和噬菌体 P22 之间具有以下平行关系而推测 F_A 因子正是 P22。

(1)P22 的侵染性和 FA 的能力都可被一些水解酶，如 DNase 和胰蛋白酶所抑制；

(2)FA 的大小和质量与 P22 相同。

(3)用 P22 抗血清或加热处理都可以使 P22 和 FA 失活，且失活速率相同；

(4)FA 和 P22 的宿主范围相同。

就这样他们发现了与转化、接合作用不同的细菌传递遗传物质的第三种途径——转导。

二、普遍转导

由于在 20 世纪 50 年代沙门氏菌的遗传学仍是未知的领域，所以转导的发现未能得到发展。1955 年 E. Lennox 发现了 P1 噬菌体可在 *E.coli* 中转导遗传物质。因 *E.coli* 研究得较多，所以 P1 的转导就可以得以发展。转导噬菌体是怎样产生的呢？1965 年 K. IKeda 和 J. Tomizawa 用 *E.coli* 的 P1 噬菌体实验回答了这个问题。

现在已知 P1 噬菌体的 DNA 为双链线状，宿主为 *E.coli*，它既有溶原周期又有裂解周期。P1 的特点是：①感染宿主后，每次可释放 100 个子代噬菌体；②其 DNA 分子量为 66×10^6 Da，约 10^5nt；(3)P1 编码一个核酸酶，能降解 DNA，但作用较慢。当噬菌体开始包装时，宿主的 DNA 大部分成为分子量为 10^7 ~ 10^8 Da 的一些片段，因 P1 包装不严

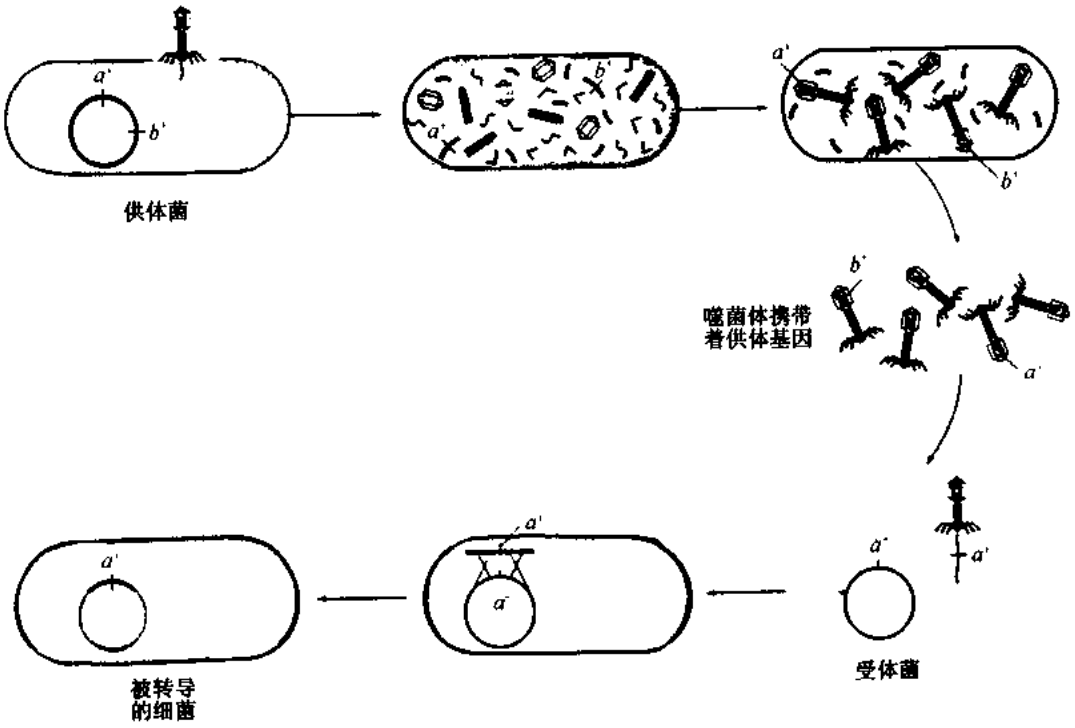
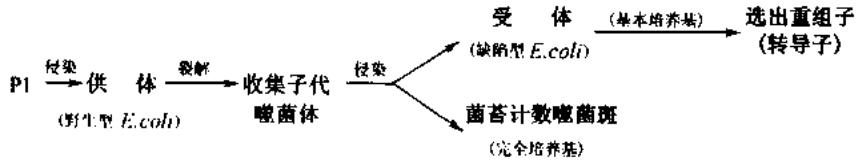


图 8-28 普遍性转导的机制。其实只有很少的噬菌体后代(1/1000)带有供体基因。

(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 10-29)

格，其中约 1/100 的宿主片段可能会被错误地包装到噬菌体中，称为"**偷渡**"。大部分转导颗粒中含有 $20\times10^6\sim25\times10^6$Da 的 DNA，外源 DNA 片段的长度只能共转导 3% 的遗传距离。转导过程如图 8－24 所示，噬菌体感染 *E.coli* a^+ 后，进行繁殖，同时使宿主的 DNA 降解成小片段，P1 噬菌体包装时有部分错误地将宿主的 DNA 包装到头部，细胞裂解后一并释放出来。当带有宿主 a^+ 基因的转导颗粒再去感染 *E.coli* a^- 时，因其不含 P1 DNA，所以不能复制、繁殖和裂解，只是将含 a^+ 的 DNA 片段射入新的 a^- 受体，经双交换发生重组使 a^- 细胞转导成 a^+，带有 a^- 的片段是线状，可被降解掉。对于任何特定的基因来说，大约以 10^{-5} 的频率形成转导颗粒。

转导频率可用以下方法来计算：



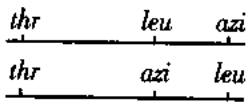
因转导颗粒不能裂解，所以不形成噬菌斑，因此侵染受体总的 P1 应为噬菌斑数加转导子数。故转导频率为：

$$\text{转导频率}=\frac{\text{转导子数}}{\text{侵染受体的 P1 颗粒数}}\times100\%=\frac{\text{转导子数}}{\text{噬菌斑数}+\text{转导子数}}\times100\%$$

三、共转导

一个以上的基因同时被转导的现象称为**共转导**(contransduction)。共转导也可用来作图。中断杂交实验方法复杂，且结果不精确，重组作图构建交互突变的品系也很不容易，而采用共转导的方法既简单又精确，同时无须构建交互突变的供体、受体品系。由于转导的片段较短，一般不能共转导 3 个基因，因此可以提供紧密连锁的资料。现以供体为 *thr*$^+$ *leu*$^+$ *azi*^{r} *E.coli*，受体为 *thr*$^-$ *leu*$^-$ *azi*^{r}，用 P1 进行转导，结果如表 8－9 所示，如果两个标记离得很近，那么共转导的可能性就大，若离得较远那么共转导的可能性就小。当 *leu*$^+$ 为选择标记时 *azi*^{r} 共转导的频率为 50%，而 *thr*$^+$ 仅为 20%，可见这三个基因中 *leu*$^+$ 离 *azi* 较近，按此它们有

两种可能的排列：



为了进一步搞清这两种排列中哪已知正确，只要知道 *thr* 离 *leu* 和 *azi* 的远近便可知晓，所以再进行第二组实验(见表 8－9)。当以 *thr*$^+$ 为选择标记时 *leu*$^+$ 的共转导频率高于 *azi*，说明 *thr* 离 *leu* 较近，也就是第一种排列是正确的。我们还可以用实验来确定一下这一结论。假设第二种排列是正确的，那么当我们选择 *thr*$^+$ 和 *leu*$^+$ 为标记时，夹在中间的标记除存在双交换之外，应以很高的频率共转导。但实验结果正好相反，*azi* 的共转率为零，说明这一假设不能成立，而只有第一种排列是正确的。

用普遍性转导来作图，常借助于 2 个或 3 个因子的转导实验，经分析转导子存在

或缺少哪些供体标记。如由 a^+b^+ 供体转导给 $a\ b$ 受体，可产生 a^+b、$a\ b^+$ 和 a^+b^+ 转导体，我们选择 a^+ 或 b^+ 就可以确定这两个基因连锁距离，公式为：

表 8-9 P1 共转导实验结果

实验	选择标记	非选择标记	
1	*leu*$^+$	*azi*^{r} ：500%	*thr*$^+$ ：2%
2	*thr*$^+$	*leu*$^+$ 3%	
3	*thr*$^+$ *leu*$^+$	*azi*^{r} ：0%	

$$RF_{ab} = \frac{\text{单个转导子数}}{\text{总转导子数}} \times 100\%$$

当我们选择标记为 a$^+$时，公式为

$$RF_{ab} = \frac{a^+b}{a^+b + a^+b^+} \times 100\%$$

当我们选择标记为 b^+时，公式为

$$RF_{ab} = \frac{ab^+\text{数}}{ab^+ + a^+b^+} \times 100\%$$

由于产生转导子是概率很低的随机事件，加上噬菌体能携带细菌的 DNA 的总量受其头部的大小容量所限制，所以无论是 a^+ 还是 b^+ 被错误地包在噬菌体头部的机会和它们的距离成正比。若一些基因距离十分近，可以同时包在转导噬菌体中的话就会产生的 a^+b^+ 共转导。两个或多个基因的共转导是紧密连锁的一种很好的指标。两个共转导基因之间的距离可以通过转导来确定，建立精细的结构图。

四、特异性转导

特异性转导又称局限性转导(restricted transduction)，是由 J. Lederberg 和他的学生所开拓的。他们设计了一个实验，目的是想了解 λ 噬菌体是否像 P22 那样将供体的基因传递给受体。结果发现了另一种机制的转导。他们将紫外线诱导野生型 *E.coli* K12 (λ) *gal*$^+$ 中的原噬菌体，使细胞裂解，收集裂解液去感染非溶原的品系，结果在基本培养基上选择到 10^{-6} *gal*$^+$ 的转导体。进一步发现：①这些转导子是溶原的，具有“免疫力”，说明噬菌体已进入受体的染色体上。②*gal*$^+$ 转导子遗传特点不稳定，有 1% ~ 10% 又变成 *gal*$^-$，表明它又失去了 λ*gal*$^+$，因此推论 λ*gal*$^+$ 是插入到受体的染色体中，而不是通过双交换而产生的。实际上既有插入，又有双交换。不稳定的部分确实由整合、环出或切离所造成。③*gal*$^+$ 转导子在细菌裂解时不能产生成熟的噬菌体，即缺失 - 取代颗粒(deletion-substitution particles)。

现已知道 λ 噬菌体和 *E.coli* 之间有一个同源的特定位点叫做**附着位点**(attachment site)，简写为 *att*。*E.coli* DNA 上的称为 *attB* (又叫 *att*λ)，由 B、O、B′三个序列组成。B 是Bacteria 的缩写；λ 噬菌体 DNA 上的称 *attP*，由 P、O、P′三个序列组成（见第 17 章），P 是Phage 的缩写。正常的 λ 染色体可通过 *att* 位点整合到 *E.coli* 的染色体中，呈原噬菌体状态（图8-29a），进入溶原周期。在紫外线的诱导或自发

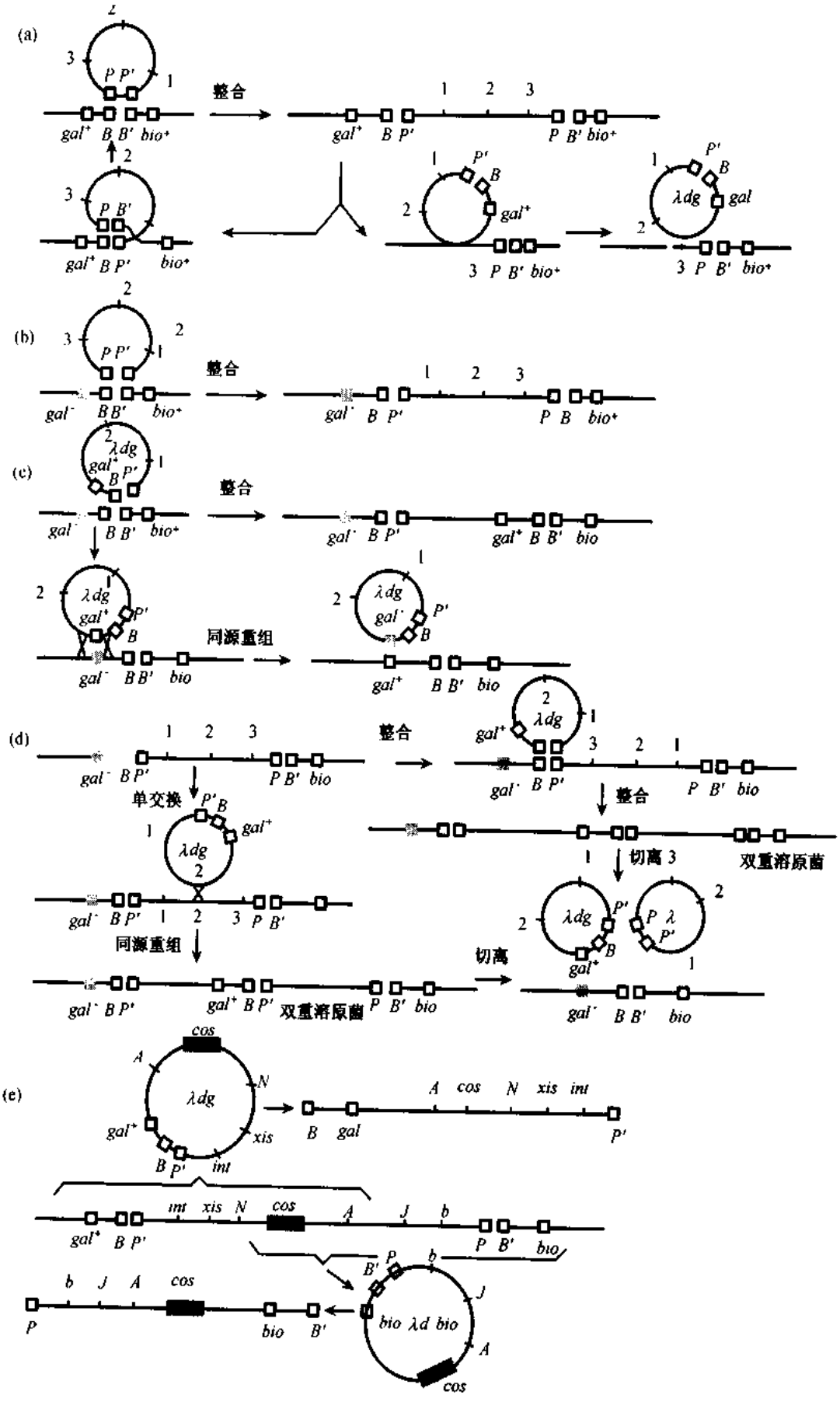


图 8－29　λ噬菌体的 LFT 和 HFT 以及 λdg，λdbio 的形成。

治愈时，λ 染色体又可以切离出来，但切离有两种情况，一种是**正常的切离**，切离出的基因组是完整的；另一种是**非正常切离**，将 *att* 位点附近的基因 *gal* 或 *bio* 错误地环出，而将本身的 DNA 部分片段遗留在 *E.coli* 的基因组中（图 8－29a 和 e）。λdg（λd*gal*）是指带有 *gal* 基因的缺陷型 λ，其尾部基因 J 丢失，不能复制；λ*pgal* 表示带有 *gal* 基因的另一种缺陷型 λ，仅丢失了 b 区，菌体仍然裂解，形成噬菌斑，“P”表示噬菌斑（*plaque*）。λ*dbio*，表示带有宿主的 *bio* 基因，而丢失了反终止蛋白基因 N，故不能形成噬菌斑；λ*pbio* 是表示带有 *bio*，而丢失了 *int* 和 *xis* 基因，这两个基因和整合有关，不影响裂解，故可以形成噬菌斑。正常的 λ 因其未携带 *gal*$^+$ 或 bio，所以它侵染 *gal*$^-$ 或 *bio*$^-$ *E.coli* 时，不能使它们转导成野生型（图8－29b）；λ*dg* 侵染 *gal*$^-$ 菌株时可以通过两种不同的途径进行转导（图8－29c）：一种是通过整合形成部分二倍体，这种转导子不稳定，可因 λ 的切离、丢失而又回复成 *gal*$^-$ 品系。另一种是通过 λ 携带的 *gal*$^+$ 和受体的基因 *gal*$^-$ 进行同源配对，经双交换，产生重组的转导子，这种转导子较为稳定。由于 λdg 多不能复制，经诱导使细胞裂解后，只能释放出 1 个 λ*dg* 去感染受体，转导的频率较低（小于 10^{-6}），故称为**低频转导**（low-frequency transduction，LFT）。

当带有 *gal*$^+$ 的 λ*dg* 感染 *E.coli gal*$^-$（λ）时，可通过两种不同途径形成**双重溶原**（double lysoge）细菌，一种是通过整合，另一种是通过 λ*dg* 和已整合的正常 λ 之间通过单交换进行同源重组所组成。双重溶原菌的染色体上含有一个正常的 λ 可提供给 λ*dg* 复制所需的一切物质，因而 λ 和 λ*dg* 同步大量复制，直至细胞裂解释放出来，λ 感染敏感性细菌可形成噬菌斑，噬菌斑数既表示正常 λ 数，也表示 λ*dg* 的总数。λ*dg* 可感染 *E.coli gal*$^-$ 受体，进行转导。由于一个细胞能释放出比 LFT 多得多的 λ*dg*，所以转导频率较高，故称为**高频转导**（high-feequency transduction，HFT）。

第九章

遗传物质的分子结构和性质

基本概念

DNA 和 RNA 大分子是由多个核苷构成，每个核苷酸又是由一个戊糖和 4 种碱基及一个磷酸基团构成的。

DNA 分子含有腺嘌呤、鸟嘌呤、胞嘧啶和胸腺嘧啶。RNA 分子含有腺嘌呤、鸟嘌呤、胞嘧啶和尿嘧啶。

DNA 分子含有两条多核苷酸链，此链以氢键相连（A═T，G≡C）形成双螺旋。双螺旋的直径是 2nm，每个螺旋有 10 个碱基对，每个碱基对的间距为 0.34nm。

双螺旋 DNA 有三种主要的不同构象，A，B 和 Z－DNA，生理条件下多呈 B－DNA 型。

双螺旋中碱基互补为半保留复制提供了前提。

复性动力学可以用来测定原核生物的分子量，鉴别重复顺序和单一顺序。不同的生物变性曲线不同。直核生物的 DNA 复性范围大于原核生物。

复性动力学为各种杂交以及 PCR 技术奠定了理论基础。

在孟德尔发现遗传法则之后不久于 1868 年瑞士青年生物化学家 J.T.Miescher 发现细胞核中含有一种富有磷的酸性物质，定名为核素（nuclein），后来称为核酸。他的这项重大发现和遗传法则的命运相同，没有受到人们的重视，直到 20 世纪 30 年代末，人们才逐渐将核酸化学的研究和细胞的功能联系起来，1944 年 O.T.Avery 等发表了“诱导肺炎球菌类型转化的物质的化学特性研究”这篇划时代的成果论文，第一个证实遗传物质是 DNA，但在这以前较流行的观点认为蛋白质是遗传物质。

第一节　核酸是遗传物质

早在 1883～1889 年间提出种质学说的 Weismann 就预言道：“遗传物质是具有特定分子结构的化合物”。在证实 DNA 和 RNA 是遗传物质之前很长的时间里，遗传学家们认识到，遗传物质这种特殊的分子必须具备以下基本特点。

（1）它必须稳定地含有关于有机体细胞结构、功能、发育和繁殖的各种信息；

（2）它必须能精确地复制，这样后代细胞才能具有和亲代细胞相同的信息；

（3）它必须能够变异。如果没有变异（如通过突变和重组）生物就不能改变、适应，进化也不会发生。

从 19 世纪中期到 20 世纪初，科学家们都认为遗传物质是蛋白质，19 世纪后期，人们在显微镜下观察到了染色体。20 世纪的前 40 年 A.Kossel，P.A.Levene 和 W.Jones 通过化学分析，发现脱氧核糖核酸（DNA）的化学组成，但却不知道 DNA 是构成基因的化学物质。先天性代谢障碍的发现以及一基因一酶学说的提出（见第 8 章）明确了基因和蛋白质的对应性，使人们考虑到基因在信息上与其产物相关。

一、遗传物质的发现

1928年英国 Frederick Griffith 的转化实验导致了遗传物质的发现。他所用的肺炎双球菌(*Streptococcus pneumoniae*)是一种光滑型(S型),在其细胞壁的外面有一层多糖的荚膜,S型双球菌感染小鼠会导致小鼠患败血症而死亡,但加热杀死S型双球菌后再感染小鼠是不会引起患病。R型肺炎双球菌的外面没有荚膜故称为**粗糙型**,R型菌感染小鼠是不会引起小鼠的死亡,但如果将R型菌和加热杀"死"的S型菌混合后感染小鼠能引起小鼠死亡(图9-1),并在其心血中检出有活的S型细菌。Griffith 认为实验的结果是由于那些加热杀死的S细菌的存在导致那些活的R细菌发生**转化作用**(transformation),从而使R型恢复了生成荚膜的能力。三年后发现只需有加热杀死的S细菌的存在,也可导致R型双球菌在体外发生转化。又过了两年,进一步证明只需对S细菌的提取液加到生长着的R型细菌培养物中,同样也能发生转化。那么提取物中究

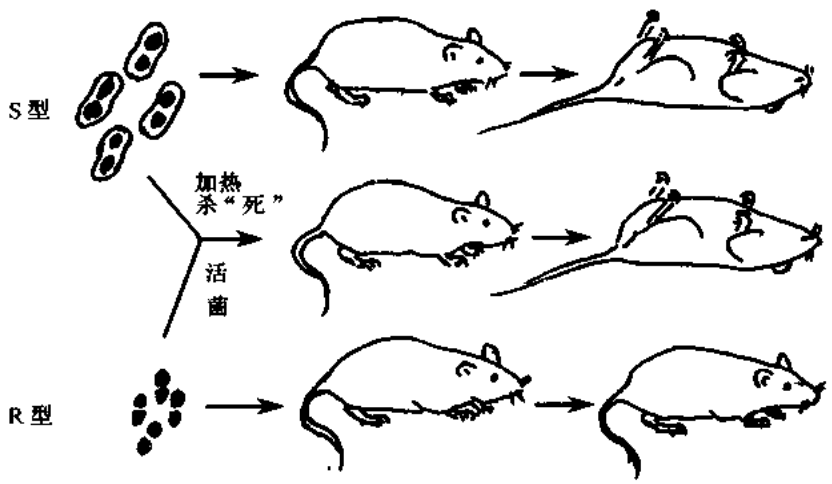


图9-1 肺炎双球菌的转化实验。

竟是什么物质使S细菌发生转化呢? Griffith 最初认为"接种物中的死细菌可能提供了某些特异的蛋白质为食料,使R型细胞能制造荚膜"还有人提出:病毒学说和感染性诱变因素等假设。Oswald Avery,C.M.Macleod 和 M.Mccarty 在前人工作的基础上,经过了十年的努力,终于完成了体外转化实验,弄清了这种转化因子的化学本质是DNA,而不是蛋白质或其他的大分子。他们将S Ⅲ 加热杀死后,分离提取多糖、脂类、RNA、蛋白质、DNA 等分别加入R菌中,培养后,仅加入S Ⅲ DNA 的R Ⅱ发生转化,产生了R和S(图9-2)。同时他们还用不同酶来处理提取物,观察对实验的影响,结果仅在DNA中加入一种使DNA降解的酶时就不能发生转化了。酶的加入不仅是一种反证,同时也说明即使在DNA不纯的情况下,比如带有少量的蛋白等,也能雄辩地证明使R Ⅱ发生转化,产生S Ⅲ 的因素惟有DNA,而不是任何别的物质。但当时人们仍不相信这一结论,这是由于:①虽然 R.Feulgen (1924年)已证明了DNA是染色体的主要组分之一,但人们仍然认为遗传和染色体上的蛋白质有关。因认为蛋白分子量大,结构复杂,其中二十种氨基酸不同的排列组合将是个天文数字,可以作为一种遗传信息,且在不同

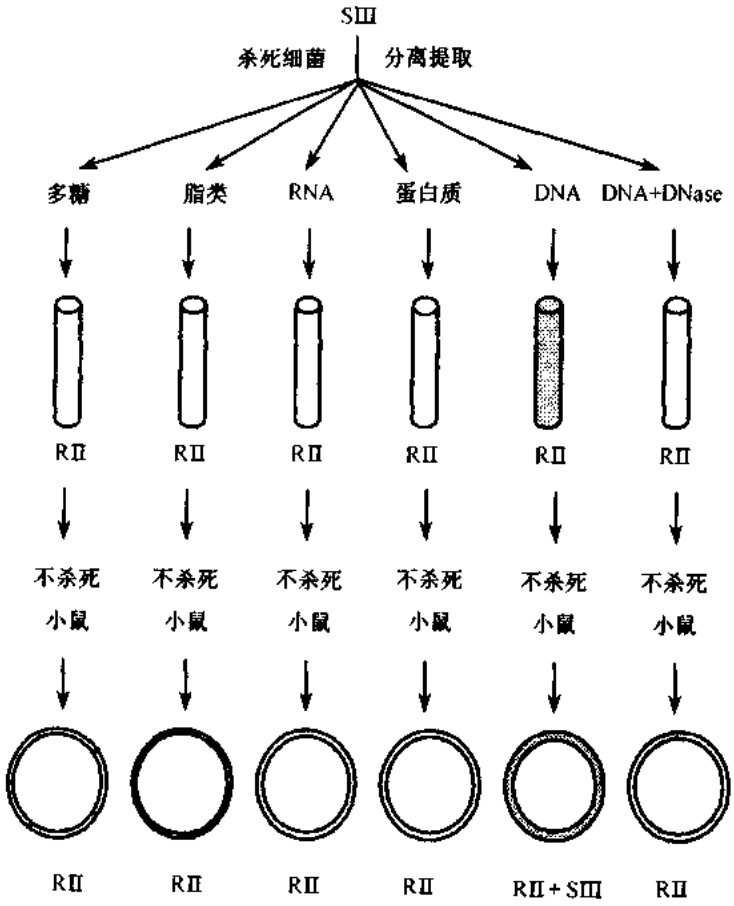


图 9-2　O.Avery 等的体外转化实验。

生物体中的同源蛋白之间在结构的特异性上存在着极大的差异。而 DNA 分子量小，只含 4 种不同的碱基，人们一度认为不同种的有机体的核酸只有微小的差异，因而形成一种观念，基因和染色体的活性成分是蛋白质。②认为转化实验中 DNA 并未能提得很纯，还附有其他物质，可能正是这些少量的特殊蛋白在起转化作用。当时人们难以忘记 20 年前著名的生化学家 Willstatter 由于不能将酶提纯而错误宣称酶不是蛋白的沉痛教训，耽心重蹈覆辙的心理增加了人们对 Avery 等实验结果的怀疑。③也有人认为即使转化因子确实是 DNA，但也可能 DNA 只是对荚膜形成起着直接的化学效应，而不是充当遗传信息的载体。由于以上原因 Avery 的重大发现却未能引起人们的重视，即使到 1949 年 Hotchkiss 证实了和荚膜无关的细菌性状也能转化，用实验证明了 DNA 已提得很纯，其中蛋白质的污染已降到 0.02%，这种纯的 DNA 仍可转化，且纯度越高转化效率也愈高，但仍未能改变人们的观点，直到 1952 年 Hershey-Chased　实验发表才使人们信服 DNA 是遗传物质。

二、Hershey-Chase 的噬菌体实验

Avery 及其同事的体外转化实验是十分精确和严密的，但很多的科学家仍不相信 DNA 是遗传物质，所幸的是在 Avery 去世（1955）前，终于在 1952 年由于 Hershey-Chase 的实验使 DNA 是遗传物质的结论得到了进一步的证实。这一突破性的实验是受到

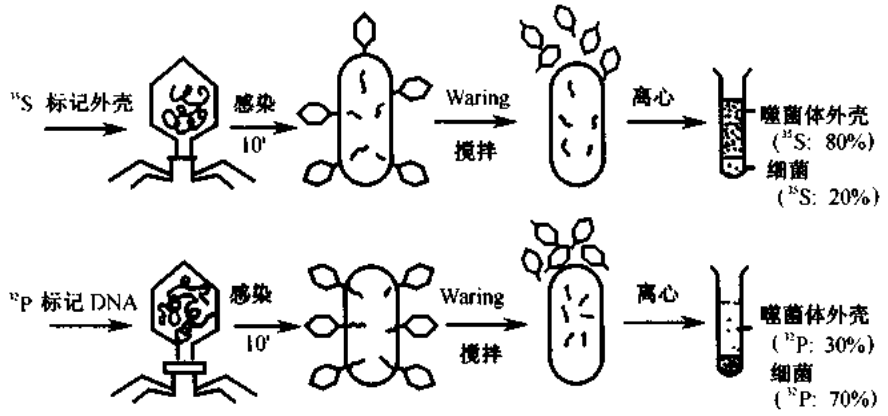


图 9-3 Hershey-Chase 的噬菌体震渗实验。

T.F.Anderson 两项发现的启发，即 1949 年 Anderson 发现将 T-偶数噬菌体悬液骤然用蒸馏水稀释，使其受到**渗震作用**时，噬菌体释放出 DNA(图 9-3)留下头部中空的噬菌体的空壳；噬菌体用尾部吸附到细菌表面上，形成动力学不稳定的联结，用组织搅碎器剧烈搅拌就可以阻碍侵染作用。Herriott（1951）发现这种释放了 DNA 的噬菌体空壳仍可吸附到细菌上。这些发现为噬菌感染实验打下了基础。当时已知 T2 噬菌体是由蛋白质外壳和内部的 DNA 组成。蛋白质中含有硫而不含磷，DNA 中含磷而不含硫，所以 Hershey 等分别用同位素^{35}S 和^{32}P ，来标记 T_2 的蛋白外壳和 DNA。他们首先将 T2 噬菌体分别感染在含有^{32}P 和^{35}S 培养基中的两组 *E.coli*，细胞裂解后分别收集裂裂菌液，经标记后再分别感染 *E.coli*，感染后培养 10 分钟，用 Waring 氏搅拌器剧烈搅拌使吸附在细胞表面上的噬菌体脱落下来，再离心分离，细菌是在沉淀中，而游离的噬菌体悬浮在上清液中。经同位素测定，上清液中^{35}S 的含量为 80%，沉淀中含量为 20%，这表明蛋白质的外壳脱落下来，并未进入细胞中，沉淀中的 20%可能由于少量的噬菌体经搅拌后，仍吸附在细胞上所致。而^{32}P 相反在沉淀中含有 70%，而在上清液中仅有 30%。表明噬菌体感染细菌后将带有^{32}P 的 DNA 已注入细胞中，可能还有少部分噬菌体尚未将 DNA 注入宿主就被搅拌了下来，所以上清液中约有 30%的^{32}P(图 9-4)。这个实验的结果，进一步证实 DNA 是遗传物质，而不是蛋白质。鉴于当时有许多研究者已开始注意到 DNA，因此和 Avery 的时代不同，他们的结论很快得到了承认，Hershey 也因此荣获了 1969 年诺贝尔医学生理奖。

三、RNA 也是遗传物质

大部分生物的遗传物质是 DNA，但也有某些病毒，如烟草花叶病毒(tobacco mosaic virus，TMV)，其遗传物质是 RNA。

烟草花叶病毒是筒状的(图 9-4)直径为 18nm，长度为 300nm。由 2130 个蛋白亚基构成外壳，环绕着一个 RNA 分子核心。把 TMV 放在水和苯酚中震荡，就可以把病毒的蛋白质部分和 RNA 分开，也可由这分开的两部分重新组合成有感染力的病毒。1956 年 A.Gierer 和 G.Schraman 发现从 TMV 分离出的 RNA 也能侵染植物，产生典型的 TMV 所致的病斑，如用 RNA 酶处理，则 RNA 就失去感染能力。而分离出的蛋白部分无这种感染力。这个实验的结果证明了 TMV 的遗传物质是 RNA，而不是蛋白质。1957 年美国的 Heinz Fraenkel-Conrat 和 B.Singre 用实验进一步证实了这一结论。他们从 TMV 中分离出

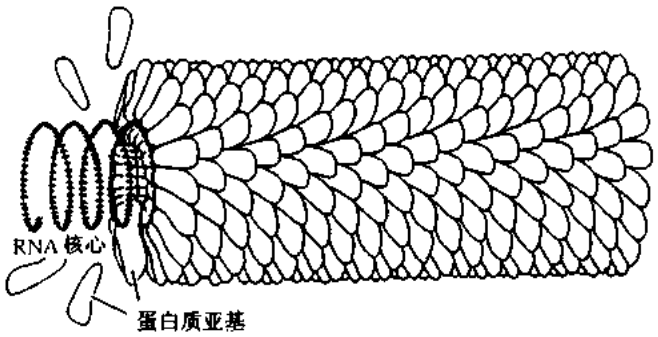


图 9－4　烟草花叶病毒的形态结构。蛋白质亚基环绕着螺旋状的 RNA 核心。
（转引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 9－8）

两个不同品系：M(masked strain) 和 HR(Holmes ribgrass strain)，它们感染植物产生的病斑形态不同（图 9－5）。若将 M 和 HR 品系重建，即将两种病毒蛋白质外壳互换一下，可重建成“杂种”TMV，它们仍具有感染能力，所产生的病斑形态和“杂种”TMV 的 RNA 一致，而和蛋白质无关。由此看来病斑的遗传信息不是由蛋白质传递的而是由 RNA 来传递的，根据以上实验可以得出结论核酸是遗传物质。

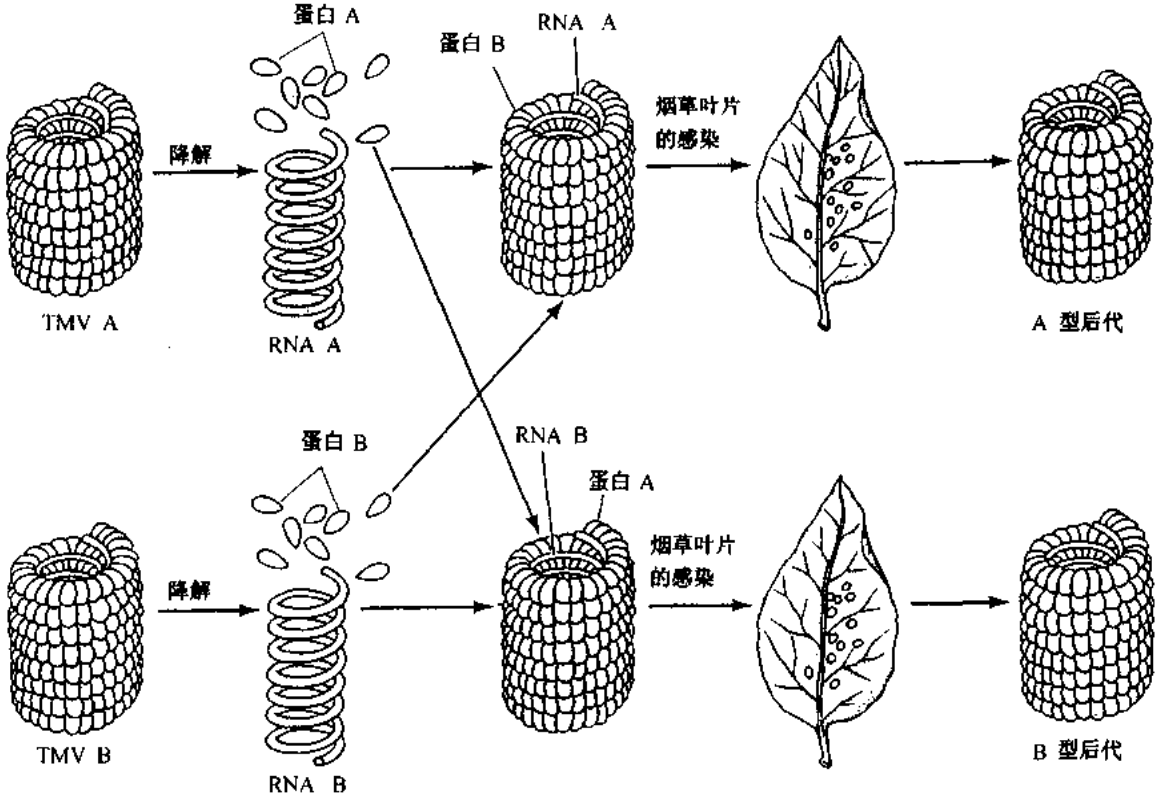


图 9－5　烟草花叶病毒的重建实验。（引自 Russell. P. J: Genetics, 3rd ed., 1992, Fig. 9－9）

第二节 DNA 和 RNA 的化学组成及双螺旋模型

一、DNA 和 RNA 的化学组成

DNA 和 RNA 是生物大分子之一，分子量至少在几千个 Da。核酸经水解成单核苷酸(nucleotides)，单核苷酸由磷酸基因(phosphate group)和核苷(nucleotide)。核苷含有戊糖(pentose)和碱基(base)。在 DNA 中戊糖是 D－2－脱氧核糖，碱基是 A、T、G 和 C。而

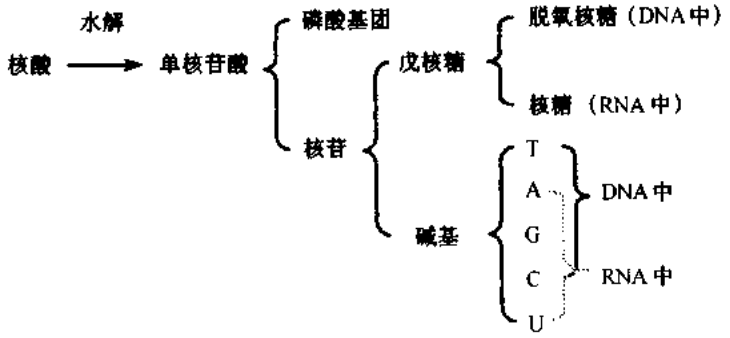


图 9－6 核糖的化学组成。

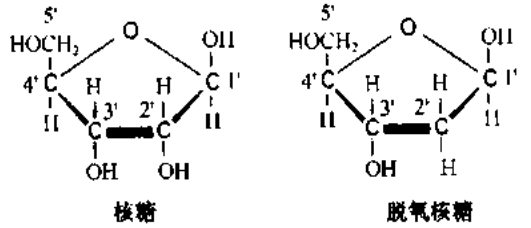


图 9－7 核糖和脱氧核糖的结构。RNA 和 DNA 的戊糖其差异在于第二个碳原子上的基团，前者是羟基，后者是氢原子。(引自 Russell, P.J:Genetics,3rd ed.,1992,Fig.9－10)

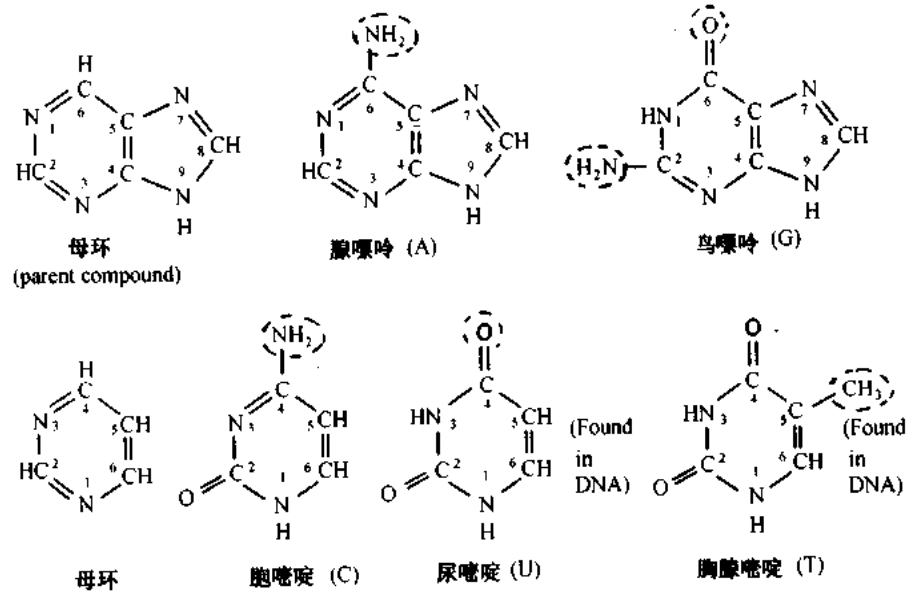


图 9－8 在 DNA 和 RNA 中含氮碱基的结构。有复合环的是嘌呤(上)，单环的是嘧啶，二者的差异是明显的。(引自 Russell,P.J:Genetics,3rd ed.,1992,Fig.9－11)

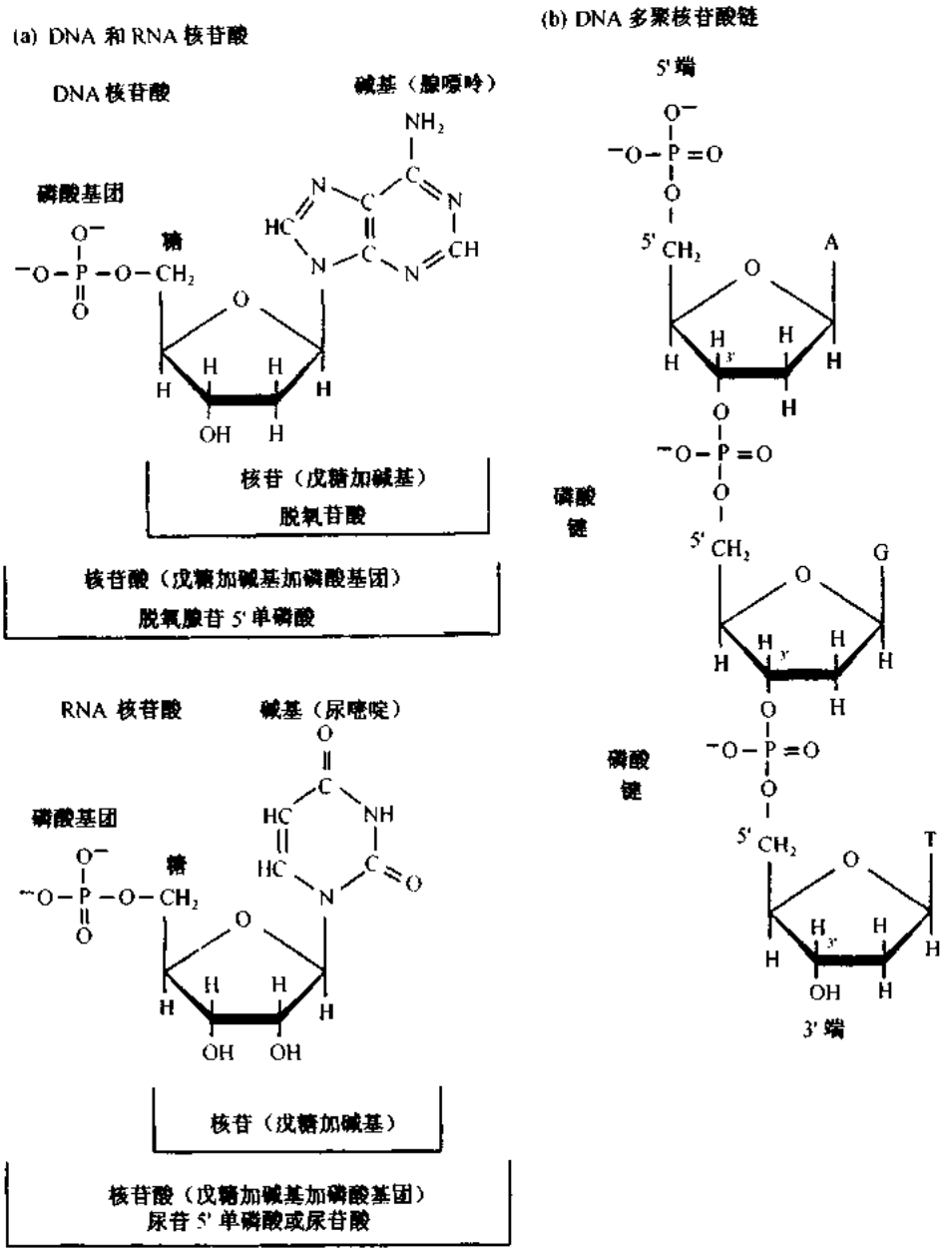


图 9－9　DNA 和 RNA 的化学结构。(a)DNA 和 RNA 的核苷(戊糖加碱基)与单核苷酸(戊糖加碱基加磷酸基团)的基本结构;(b)DNA 的多核苷酸单链。脱氧核糖的 3′碳原子和下一个脱氧核糖的 5′碳原子之间通过磷酸键相连。(引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 9－12)

RNA 戊糖是 D－核糖,碱基是 A、U、G 和 C(图 9－6～9)。

起初人们普遍认为 DNA 是由 4 种核苷酸有规律地重复排列而成。DNA 分子中有等摩尔数的 A、T、G 和 C,即 DNA 的四核苷酸假说:不论其来源如何,DNA 的成分都是相同的,因此不可能作为遗传信息的载体。1950 年英国的 Chargaff 应用先进的纸层析及紫外分光光度计对各种生物 DNA 的碱基组成进行了定量测定,他发现了一个重要的规律:"尽管沿着多核苷酸链的碱基是复杂的,不同的 DNA 其碱基组成显著不同,但腺嘌呤和胞嘧啶的数目总是相等的。"此表明:①所有 DNA 中 A 与 T 的摩尔含量相等,

即[A]=[T],G与C的摩尔含量也相等,即[G]=[C]。因此嘌呤的总含量和嘧啶的总含量是相等的A+G=C+T,这一规律称为"当量规律"。②不同生物的DNA碱基组成有明显的差异,表明具有物种的特异性,但没有组织的特异性和不同发育阶段的特异性,即各种生物的碱基组成是特异而稳定的,否定了四核苷酸假设。Chargaff对这一发现谨慎地猜测为"这是DNA结构的某些重要面貌的反映"。在双螺旋模型未诞生之前,能有这样的"猜测"是有远见卓识的。实际上各种生物DNA碱基组成的特异性和稳定性表明DNA是可以携带遗传信息。当量规律的发现提示了A与T,G与C互补的可能性,为双螺旋模型的建立提供了重要依据。不久以后Alfred Mirsky, Hars Ris和R.Vendrely, A.Boivin两组学者分别发现不同的生物体细胞中DNA的含量都是其配子中的两倍,此和染色体遗传学的原则完全一致。此一发现扭转了局面,使得DNA是遗传物质的观点迅速成为主流。

二、DNA双螺旋模型的诞生

1951年美国微生物学家S.Luria的博士研究生J. D.watson去英国进修生物化学,Watson在芝加哥大学读本科时兴趣主要在鸟类,到了高年级时产生了另一种想法:想了解基因是什么,1949年他带着这种想法进入了剑桥大学卡文迪什实验室医学研究组,与学物理出身的青年学者Crick合作,决定研究DNA的分子结构。Crick在1946年读了薛定鄂(E. Schrödinger)的名著《生命是什么》一书后,受到鼓舞,舍弃物理学而转向生命科学领域。刚到剑桥时Watson由于自己的化学与物理学的基础较差而耽心听不懂R.Franklin所做的关于DNA X-衍射研究的学术报告,而在两年之后(1953年)他与Crick却建立了DNA分子的双螺旋模型,论文虽短,却是划时代的发现。可以说是孟德尔定律问世以后,在遗传学,分子生物学领域中最为伟大的发现,因此人们将这一发现看成是分子遗传学的起点。没有双螺旋模型,就没有现在的基因工程,就没分子生物学的迅猛发展。这一模型的产生是生物学和物理学,结构化学交叉融合的结果,不同领域的科学家思维交流的结果。尽管R.Franklin和Wilkins在DNA的X衍射研究方面做得很出色,但却未能领先提出DNA结构的模型,就因为他们缺乏了生物化学方面的知识,而没有他们出色的工作,Watson和Crick所建立的模型也不能及时得到验证。

双螺旋的建立并非是科学家们脑海中的一个闪念,依然是建立在许多前人努力研究所取得成果的基础上。主要是受到4个方面的影响:①1938年W.T.Astbury和Bell用X衍射技术研究Hammorsten和Caspersson所提供的DNA。1947年拍摄了第一张DNA的衍射照片,并推断DNA分子的结构是:柱状;多核苷酸是一叠扁平的核苷酸;核酸残基取向和分子长轴垂直,间距为3.4Å。这已构成了DNA分子结构的雏形。②1951年Pauling和Corey在美国科学院报上连载了七篇有关α-螺旋结构的文章,轰动整个世界,引起了Watson和剑桥科学家们的注意,他们运用化学的简单定律来推理,而并不做具体的实验,此对Watson产生了巨大的影响。③美国晶体学者J. Donohue的指正和Chargaff的当量规律都帮助Watson-Crick纠正了起初A-A,G-G,C-C,T-T同类配对的错误想法,而提出碱基互补的正确构型。④R.Franklin和Wilkins在1952年底拍得了一张十分清晰的DNA结晶X衍射照片(图9-10);为双螺旋结构模型提供了强有力的依据和佐证。他们把数据整理好后和"DNA双螺旋结构"一文同时发表,因此,J.K.Watson、Crick和Wilkins分享了1962年诺贝尔化学奖,遗憾的是立下了汗马功劳的R. Franklin于1958年不幸英年早折,虽然她未能享受科学界的最高荣誉,可是她的卓越贡献永垂

青史。Pauling 是当时世界上第一流的理论化学家，在大分子结构的研究中屡建奇功，曾获诺贝尔化学奖和和平奖。但在 DNA 结构的研究上却败在年轻学子 Watson 的手下，错误地提出了 DNA 的三螺旋结构。真可谓智者千虑，必有一失。但给我们一个重要的启示，不同学科的交叉融合对边缘学科的突破，可能更为有利。

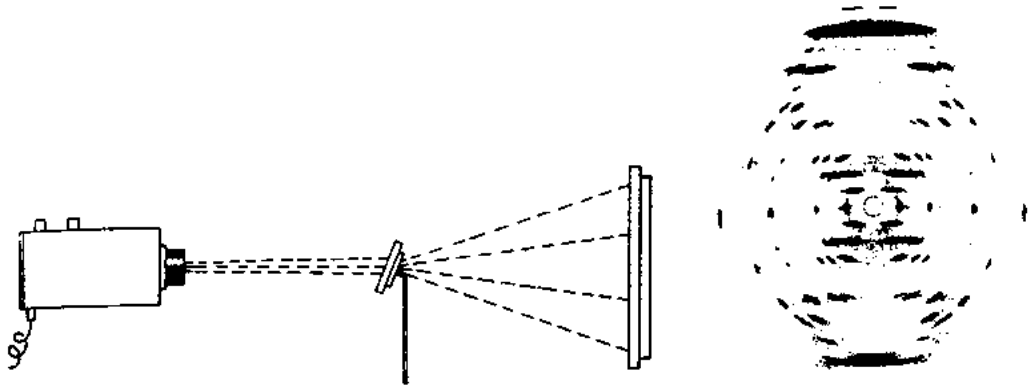


图 9－10　DNA 的 X 射线衍射分析。（引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 9－14）

三、双螺旋模型

Watson 和 Crick 的 DNA 双螺旋模型有以下特点：

（1）DNA 分子由两条反向平行的多核苷酸链组成，形成右手双螺旋，即从一端看过去，是以顺时针方向旋转。

（2）双螺旋的直径是 2nm；螺距为 3.4nm，上下相邻碱基的垂直距离为 0.34nm，交角为 36°，每个螺旋有 10 个碱基对。

（3）两条链反向平行，即两条链的 5′和 3′的方向相反。

（4）糖－磷酸键是在双螺旋的外侧，两条链上的碱基通过氢键形成碱基对，使两条链连在一起，碱基对的平面与轴线垂直。

（5）糖与附着在糖上的碱基近于垂直。

（6）碱基配对时，必须一个是嘌呤，另一个是嘧啶。

（7）DNA 双螺旋有大沟（major or wide groove）和小沟（minor or narrow groove）的存在（图 9－11～13）。

模型中的碱基配对十分重要：①A－T 和 G－C 配对时可形成很好的线性氢键；②A－T 对和 G－C 对的几何形状一样，使双链中两侧去氧核糖的 C1－C2 之间的距离相近，故不会使双螺旋丧失对称性和产生畸形；③碱基对处在同一平面内。在平面内有对称轴，垂直于平面。碱基对围绕对称轴旋 180°仍保持原来的对称性。这就是说在一个螺旋中虽碱基对在平面上的投影位置各不相同，但每对都可保持在一个平面内。那么不论核苷酸的顺序如何，都不会影响双螺旋的结构。因此不同碱基组成的 DNA 双螺旋，其外形都是光滑均匀的圆柱体；④为 DNA 半保留复制奠定了基础。

DNA 分子上有大小沟之分，这是由于碱基和核糖的连接的糖苷键同在一侧，而不是呈旋转对称，分居两侧。碱基对的糖苷键将平面分割成两个扇形区域，一个大于 180°是大沟，一个小于 180°是小沟。大沟常是多种 DNA 结合蛋白所处的空间。

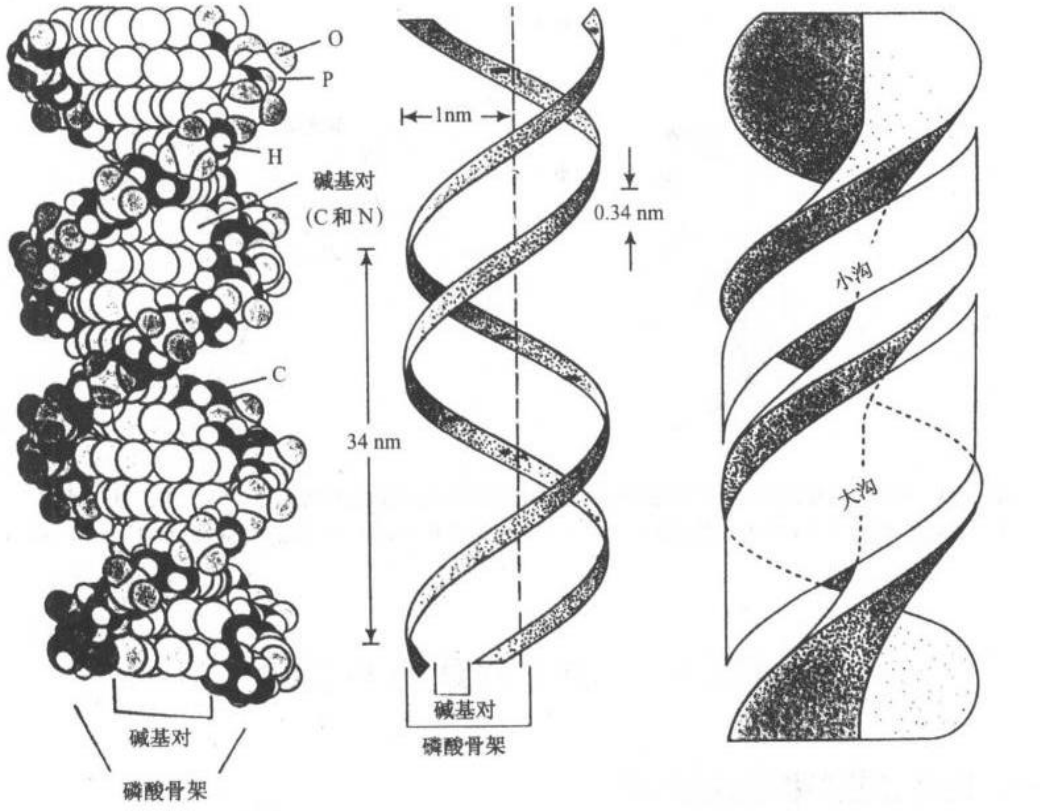


图 9-11　DNA 的分子结构。(a)DNA 的三维分子模型;(b)DNA 的双螺旋模型;(c)表示 DNA 双螺旋中的大沟和小沟。(引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 9-15)

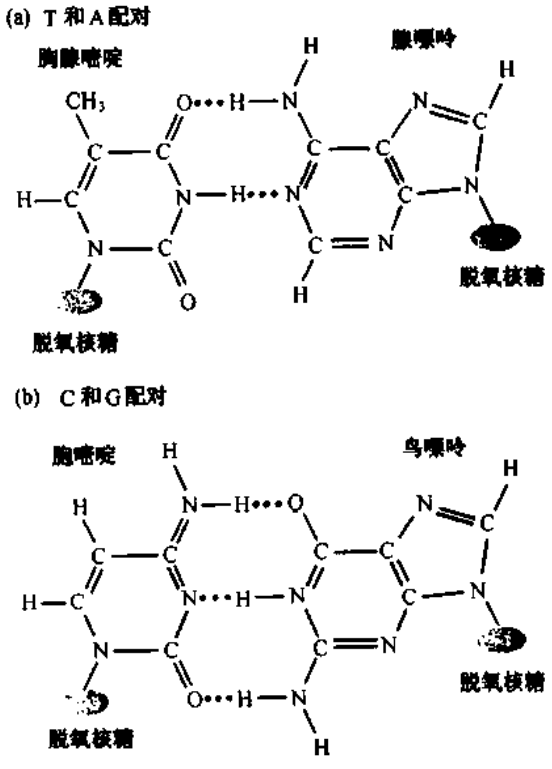


图 9-12　DNA 中碱互补配对的结构
嘌呤和嘧啶以两种方式配对:(a) T 和 A 配对;(b) C 和 G 配对。(引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 9-16)

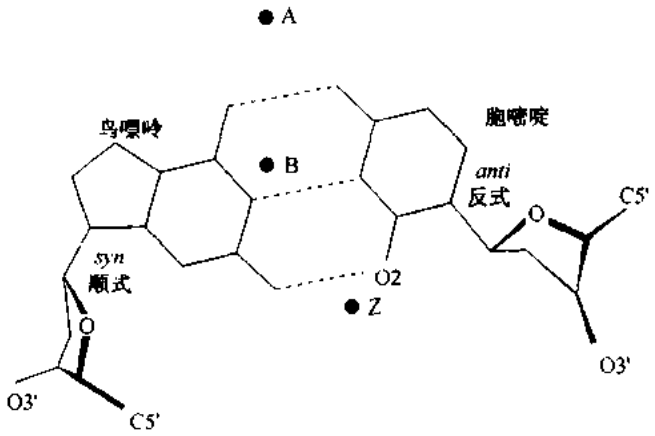


图 9－13　碱基和糖环连接的糖苷键有两种构象。右边表示反式，左边表示顺式。上方为大沟，下方为小沟。A－DNA、B－DNA 和 Z－DNA 双螺旋的中轴各不相同。（引自 Russell, P. J: Genetics. 3rd ed., 1992, Fig. 9－20）

第三节　DNA 的结构和性质

一、DNA 二级结构的稳定因素

DNA 的二级结构是很稳定的，这是由以下的因素所构成。

（1）碱基对之间的氢键。在双螺旋中嘧啶和嘌呤之间的距离正好与一般氢键的键长（0.27nm）相一致。且供体、氢原子和受体原子处于一直线上，利于形成氢键。若嘌呤与嘌呤或嘧啶与嘧啶配对，其空间的大小都不适合氢键的形成。单个氢键仍是不稳定的，分子平均的键能仅 2.5kJ/ mol，但由 20 个以上的碱基对组成的双螺旋在室温下已相当稳定。由于这个原因我们在设计 PCR 引物时，一般至少在 16nt，以使其较为稳定，但最好在 20～24 个 nt，以增加特异性。氢键的形成有助于 DNA 复制、修复、重组、转录、翻译，以及各种分子杂交等。

（2）碱基的堆集力。此是碱基对之间在垂直方向上的相互作用所产生的力。它包括①疏水作用；②范德华力；③磷酸基的负电荷斥力。DNA 的主链部分是亲水的，碱基的外围基团氨基和酮基也是亲水的，但嘧啶和嘌呤本身带有一定程度的疏水性。因此在水溶液中，这些疏水基团自发聚集。从热力学的角度来看，DNA 倾向于形成三维结构是有利于高溶性的磷酸基因与水的接触增加到最大限度，使碱基与水的接触减少到最小限度，从而使双螺旋内部相邻碱基相互聚集，形成双螺旋。在双螺旋中相邻碱基对的间隔是 0.34nm，而范德华（van der Waals）力的半径（指引力和斥力正好相平衡的距离）平均为 0.17nm。碱基间相互作用的强度与相邻碱基之间环的重叠面积成正比。总的趋势是嘌呤与嘌呤之间 > 嘌呤与嘧啶之间 > 嘧啶与嘧啶之间。

$\overrightarrow{GC}$
$\underset{\leftarrow}{CG}$ = －3.4(kcal/mol)，ΔG = －200kJ/mol；

$\overrightarrow{CG}$
$\underset{\leftarrow}{GC}$ = －2.0(kcal/mol)，ΔG = －9kJ/mol。

三个终止密码子(UAA,UAG,UGA)中,UAA 最有效。这是因为在 64 个密码子中,假如 $\overrightarrow{UAA}$ 都反密码子的话,$\underleftarrow{AUU}$ 的 Tm 最低,在生理温度下也不稳定,故也很少产生抑制基因,UAA 被自然界选择为终止密码子在热力学上是十分合理的。在碱基堆集中,处于 DNA 中间的碱基比两端的稳定,两端碱基越多中间的堆集力越大,正因为这个原因,DNA 双链两端经常绽裂(fraying)。

核苷酸的磷酸基因上都带有负电荷,因此双链之间存在着静电的排斥作用。当在溶液中加入盐类时,正离子(Na^+)可以和负电荷结合,围绕在磷酸基因的周围形成离子云,从而屏蔽了静电斥力。在处于生理盐水浓度(0.02mol/L)时就可发生这种屏蔽作用,斥力被中和,那么两条链就不易分开。当离子浓度降低,屏蔽作用减弱时,排斥力加大,Tm 值就会降低(见本章)。若浓度大于生理盐水浓度,随浓度的增加,Tm 值增加。此是因为在高盐的浓度下,碱基的溶解性降低,而增加了疏水作用,促进了双螺旋的稳定。

二、双螺旋结构的构象变异

前面介绍的双螺旋模型仅是 DNA 双螺旋多种构象中的一种,称为 **B－DNA**,存在于生理条件下,在一定的条件下双链 DNA 可以从 B 型转变成其他的构象。双螺旋的构象现已知有 A、B、C、D、E、T、Z 7 种(表 9－1,图 9－14)。引起 DNA 双链构象改变有以下因素:①核苷酸顺序;②碱基组成;③盐的种类;④相对湿度。

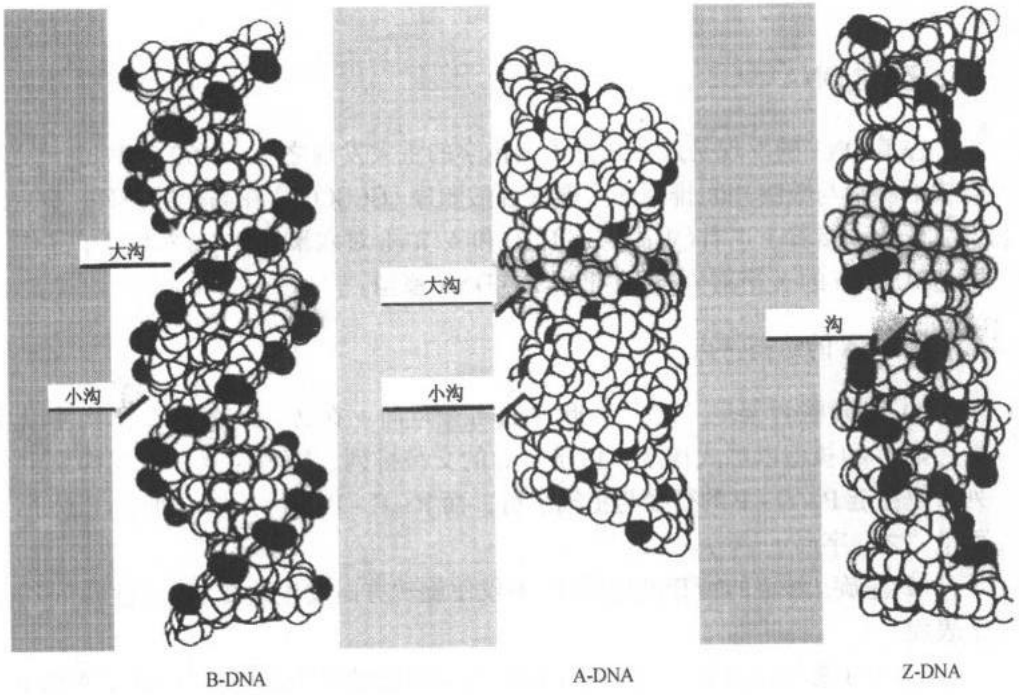


图 9－14　DNA 三种主要的构象。(引自 B. Lewin: Genes Ⅵ, 1997, Fig. 5－8)

表 9-1　A，B，C，Z 型双螺旋的特性

双螺旋类型	每螺旋内碱基对数	每对碱基的转角	每碱基对的间距(Å)	直径(mm)	存在的条件		沟型		糖苷键的构形	中轴位置
					相对温度	盐的种类	大沟	小沟		
A	11	34.7°（右旋）	2.56	2.3	75%	Na^+，K^+，Cs^+	窄深	宽深	anti(C, C)C3′ endo	大沟
B	10	34.0°（右旋）	3.38	1.9	92%	Na^+ 低盐	宽中等深	窄中等深	anti(G, C)C2′ endo	碱基对
C	9.33	38.60°（右旋）	3.32	1.9	66%	Li^+	宽中等深	窄中等深	anti(G, C)C3′ -exo	碱基对
Z	12	-30°（左旋）	3.71	1.8	43%	Na^+ Mg^{2+} 高盐	平浅	窄深	anti(G), syn(G) C2′endo C3′endo（交替）	小沟

注：C_2'或C_3'与C_3'同向称内式（endo）构象，若反向称外式(exo)构象。核苷中的碱平面绕糖苷键转动，其投影和核糖的$C_1'-O$键形成的夹角 X 为核苷的扭转角，嘌呤的扭转角 X 为C_4-N_9投影和$C_1'-O$之间的夹角；嘧啶的扭转角 X 为C_2-N_1的投影和$C_1'-O$之间的夹角　X = -60°~0~120°时为顺式构象，X = 120°~180°~-60°时为反式构象。

三、左旋 DNA

左旋 DNA 的发现是近年来分子遗传学的重大发现之一。1972 年 Pohl 等发现人工合成的嘌呤与嘧啶相间排列的多聚核苷酸顺序（GCGCGC）在高盐的条件下，旋光性会发生改变。接着王惠君(Wang，A.H.J)和 A.Rich 对六聚体 d(CGCGCG)单晶作了分辨率达 0.09nm 的 X 衍射分析，提出了 **Z-DNA 模型**。

1. Z-DNA 的结构特点

（1）糖磷酸骨架呈“之”字形（zigzag）走向。在 Z-DNA 中 C 和核糖以反式连接，而 G 和核糖以顺式连接，由于 G C 的交替排列，随着顺式，反式构象也交替排列，使主链 P-G-P 的行径与主轴平行，而 P-C-P 的行径与主轴垂直（图 9-19）而呈“之”字形。

（2）左旋。因 G 的糖苷键呈顺式，不仅使旋转方向发生改变。而且使 G 残基位于分子表面。

（3）分子外形呈波形。这是由于碱基对向螺旋的外表面前移，中轴不再像 B-DNA 位于碱基对，而移向小沟。

（4）C 的糖环C_2'为内式（endo），碱基为反式（anti），使糖环转离小沟，而 G 的糖环C_3'为内式，碱基为顺式，使糖环弯向小沟。胞嘧啶的C_5和鸟嘌呤的N_7C_8原子填满了大沟，并凸向表面，使大沟平浅，而小沟窄而深。什么是“内式”和“外

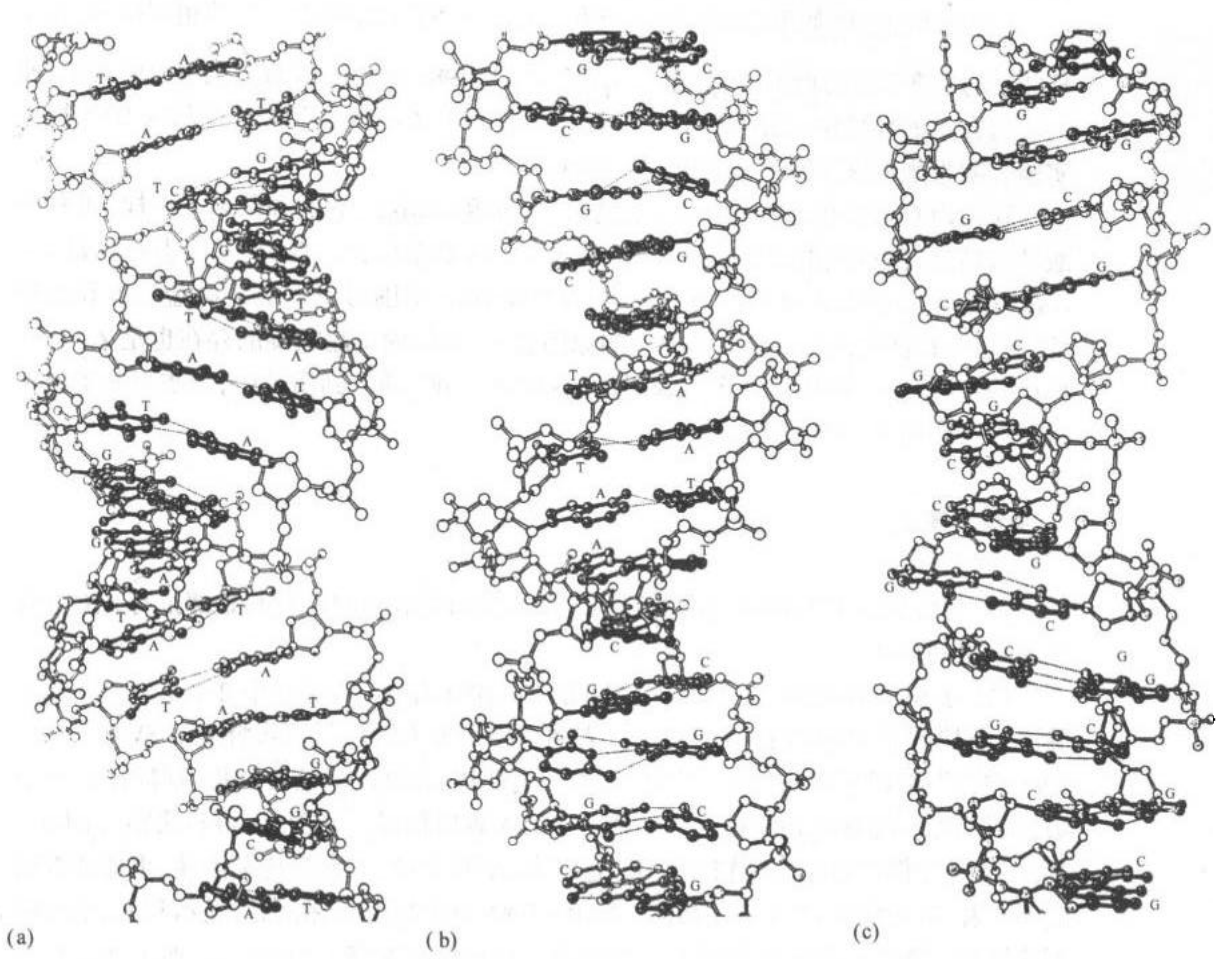


图 9－15　A－DNA，B－DNA 和 Z－DNA 的分子结构模型。灰色表示碱基对。注意在 Z－DNA 中磷酸骨架呈之字型走向；鸟嘌呤伸展到双螺旋的表面。（引自 Russell，P. J：Genetics，3rd ed.，1992，Fig. 9－19）

式”呢？戊糖的环并不是一个平面，$C_1'-O-C_4'$ 一般在一个平面上，C_2'和 C_3'常偏离平面约在 0.05～0.06nm。C_3'是在平面的上方，当 C_2' 或 C_3'与 C_5'同向时称内式构象，当与 C_5'反向时，称为外式构象。什么是顺式(Syn)和反式呢？嘌呤或嘧啶通过糖苷键和戊糖连接，它们的平面几乎和戊糖平面垂直，并可以糖苷键为轴旋转。Pu 的 C_4N_9 之间的键在戊糖平面上的投影和戊糖 $C_1'-O$ 之间的键形成夹角。同样 Py 的 C_2-N_1 的投影也可和戊糖 $C_4'-O$ 之间形成夹角，此夹角就称为扭转角。当扭转角 $\theta=0°\pm90°$时称为顺式，$\theta=180°\pm90°$.

2．Z－DNA 存在的条件

(1) 在高盐的条件下：NaCl 的浓度超过 2mol/L，$MgCl_2$ 的浓度要超过 0.7mol/L。

(2)嘌呤、嘧啶相间排列：poly d$(\frac{GC}{CG})_n$。在一定的条件下 poly d$(\frac{AC}{TG})_n$ 甚至 d$(\frac{AT}{TA})_n$ 以及更为复杂的排列如 CGCATGCG/ GCGTACGC 也可形成 Z－DNA。现在认为在适当的

离子存在条件下，任何不少于六个 bp 的嘌呤－嘧啶交替排列顺序都能形成 Z－DNA。

(3)在活细胞中如果胞嘧啶被甲基化的话(^{m5}C)则无需嘌呤－嘧啶相间排列，在生理盐水的浓度下就可产生 Z 型。如$^{m5CGAT\,m5C}_{Gm5CTAGm5C}$。1981 年 Behe 发现此是由于甲基(m)伸向大沟含水的环境中，周围局部形成憎水区，有利于 Z－DNA 的稳定。用 Z－DNA 抗体实验表明果蝇的 X 染体中就存在 Z－DNA。

在体内有多胺化合物的存在，如精胺、亚胺和亚精胺，它们和阳离子一样，可和磷酸基因结合，减少负电荷的排斥作用，使 B－DNA 转变成 Z－DNA。某些蛋白质如 Z－DNA 结合蛋白带有正电荷，与 DNA 结合可使 DNA 周围形成局部的高盐浓度和微环境，这也是在活细胞中形成 Z－DNA 的原因之一。在体内负超螺旋的存在也是 Z－DNA 形成的条件之一。抗体结合实验表明，将 poly(dG－dC)插入到质粒中，当质粒处于松弛状态时不能和 Z－DNA 抗体结合。

3. 生物学意义

(1) 可能提供某些调节蛋白的识别。啮齿类动物病毒的复制起始部位有 d(GC)有交替顺序的存在；

(2) 在 SV40 的增强子中有三段 8bp 的 Z－DNA 存在，若将其中 2 个 Z－DNA 片段除去，再接到 β－珠蛋白基因上表达，表明增强子失去活性。当将野生型 SV40 的 Z－DNA 中“T”和“C”转换成“C”、“T”时，不影响突变体的活性，但嘧啶转换成嘌呤时，由于破坏了嘌呤－嘧啶的相间排列，而使 Z－DNA 难以形成，从而 SV40 也失活（图 9－16)。另一个例子是原生动物纤毛虫，它有大、小两个核，大核有转录活性，小核和繁殖有关。大、小两核的 DNA 浓度相同。用 B－DNA 抗体以荧光标记后，显示和大、小两核都能结合，而 Z－DNA 抗体以荧光标记后，显示仅和大核 DNA 结合，而不和小核的

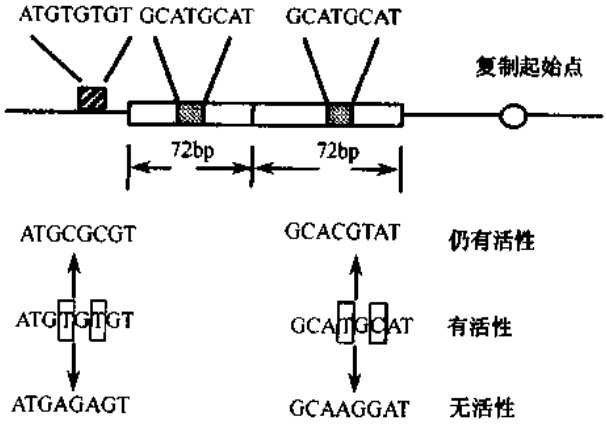


图 9－16　SV40 增强子中的 Z－DNA 及其突变后的影响。

DNA 结合，说明大核 DNA 有 Z－DNA 的存在，可能和转录有关。

四、超螺旋

E.coli 的 DNA 的总长度是其细胞长度的 100 倍，那么它的 DNA 怎样包装成类核(nucleoid)呢？原来其 DNA 存在着超螺旋。1963 年 Jevome Vinogred 在离心分离多瘤病毒

(polyoma virus) 的环状 DNA 时，原以为在离心管中只会出现一条带，结果出现了两条带，他认为一条可能是**松弛型 DNA**，另一条可能是**超螺旋 DNA**，从而发现了 DNA 的超螺旋。双螺旋进一步卷曲就会形成三级结构——超螺旋(supercoiled)。超螺旋又分成**负超螺旋**和**正超螺旋**两种。自然状态的共价闭合环状 DNA(cccDNA) 一般都呈负超螺旋状态，超螺旋的部分呈右手螺旋。某些属平面芳香族分子的药物或染料，如吖啶橙、溴化乙锭等可以插入相邻的两个碱基之间，促进产生正超螺旋，其螺旋部分是左手螺旋。闭合环若被切开一条单链，或在双链上交错切割，都会形成**开环**(Ⅱ型 DNA)，此时呈**松弛型**(relaxed form)。若此时再将缺口封闭起来，形成环状(Ⅰ型)仍是松弛形，无超螺旋存在。毫无疑问线状的 DNA 是不可能形成超螺旋的。不同类型的 DNA 的沉降系数和泳动率不同(表 9-2)，我们可以用密度离心和凝胶电泳来进行鉴别。另外也可以用电镜来直接观察(图 9-17)。

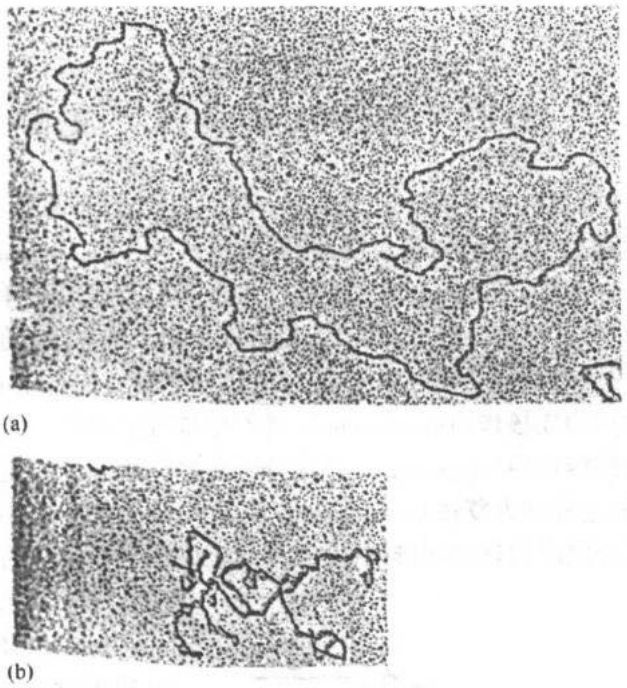


图 9-17　质粒 DNA 的电镜照片。(a)松弛型 DNA；(b)超螺旋 DNA。两个分子的放大率相同。
(引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 10-2)

表 9-2　不同类型 DNA 的三级结构及性质

DNA类型	形态	拓扑结构	沉淀系数	电泳迁移率
Ⅰ	闭合环	超螺旋	1.41	最快
Ⅰ°	闭合环	松弛形	1.41	最慢
Ⅱ	开环	松弛形	1.14	同Ⅰ°型
Ⅲ	线形	松弛形	1.00	中等速度

超螺旋是可以定量描叙的。White 在 1969 年通过微分几何的研究，建立了 White 方程(L= T+ W)来说明环绕数和超螺旋的关系。

L(Linking number)：**链环数或称拓扑环绕数**，指 cccDNA 中一条链绕另一条链的总

次数。其特点是①L 是整数；②在 cccDNA 中任何拓扑学状态中其值保持不变；③右手螺旋对 L 取正值。

T(Twisting number)：**缠绕数**，即双螺旋的圈数。其特点是①可以是非整数；②是变量；③右手螺旋时 T 为正值。

W(Writhing number)：**扭曲数**，即超数旋数。其特点是①可以是非整数；②是变量；③右手螺旋时，W 取负值。

无论是 L 数的增加或减少，即加旋或解旋，DNA 分子具有维持 B－DNA 结构的倾向，即由 10 碱基构成一个螺旋，因此，DNA 的长度确定后，“T”会保持不变，加旋和解旋的影响由三极结构的扭曲来抵消。现举例说明这三者之间的关系。如有松弛的 B－DNA，共有 420bp，一条链绕另一链 42 次，正好形成 42 圈螺旋，则 L＝T＝42，所以 W＝0，无超螺旋。若固定一端，另一端按顺时针方向旋转 6 圈，使双螺旋解开 6 圈，再将双链连接成闭合环，而 DNA 仍要保持 B－DNA 的结构，每个螺旋由 10 个 bp 组成，T 又变成最初的 42，即有 42 个螺旋，则 L＝36，T＝42，所以 W＝－6，即形成了 6 个负超螺旋（右旋），在此情况下，T 值保持不变，$\Delta T = 0$，$\Delta L = \Delta W$

第四节　DNA 的变性与复性

DNA 双链依靠氢键彼此互补地连接在一起。A－T 对有两个氢键，G－C 对有三个氢键，如果缓慢地加温，就可以使氢键断裂，双链解开，产生单链的 DNA 分子，这一过程就称为 DNA 的**变性**(denaturation)或**解链**(melting)。已经变性的单链 DNA 分子若骤然冷却，仍保持为单链，但若缓慢冷却的话，单链就重新彼此靠近，随机碰撞。若是非对应段的话即使有暂短的结合又会发生分离。互补的两条链首先是中心部分互补碱基对之间形成氢键，此称为**成核作用**(nucleation)，接着像拉拉链一样，其余部分随之形成互补双链，此称为**拉拉链作用**(zippering)。核酸分子在变性后分开的互补链缓慢冷却，重新形成互补双链的过程称为**复性**(renaturation)或**退火**(annealing)。DNA 的复性对片段有两个要求：①互补顺序的碰撞和排列；②碱基的正确配对和氢键的形成，但不需要完全配对，即不需要两条链完全互补。Roy Britten 和 David Kohne(1968 年)通过对 DNA 复性动力学的研究为我们提供了基因组(genome)中顺序分布的信息，指出了原核生物和真核生物基因组明显不同之处。

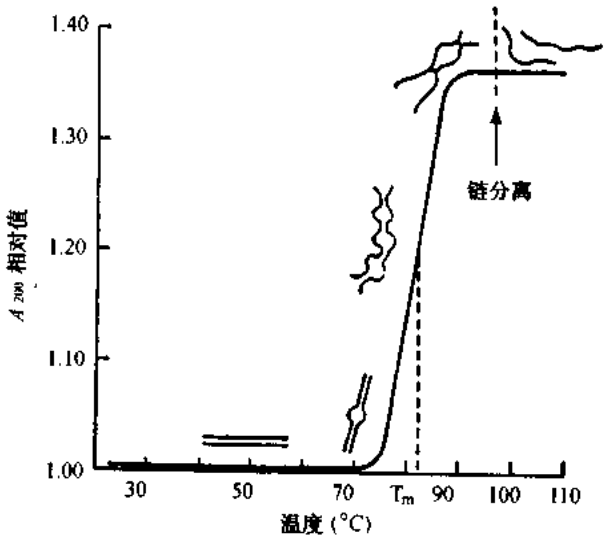


图 9－18　DNA 的解链曲线。DNA 在不同温度范围内可能呈现的不同构象。(引自 D. Freifelder: Molecular Biology, 1983, Fig. 4－6)

高温、酸、碱及某些变性剂如尿素、甲酰胺和甲酸等都可导致 DNA 变性。DNA 变性后形成单链，其物理性质也相应发生了变化：①流体力学的性质发生改变：粘度下降，而沉降速度增加；②由于发色基团藏在 DNA 的内部，变性后裸露出来，从而提高了对紫外线的吸收能力，此称为**增色效应**(hyperchromic effect)。如双链 DNA 的 $A_{260} = 1.00$

(浓度为 50μg/ml 时,对波长 260nm 紫外线的吸收能力);单链 DNA 的 A_{260} = 1.37;游离碱基或核苷酸的 A_{260} = 1.60。表示变性特征的简便参数是**解链温度**(melting temperature, T_m)或熔点,T_m 是 A_{260}的升高达到极大值一半时的温度。或者说是变性温度范围的中点。在生理条件下 T_m 一般在 85°~95°(图 9-18)。

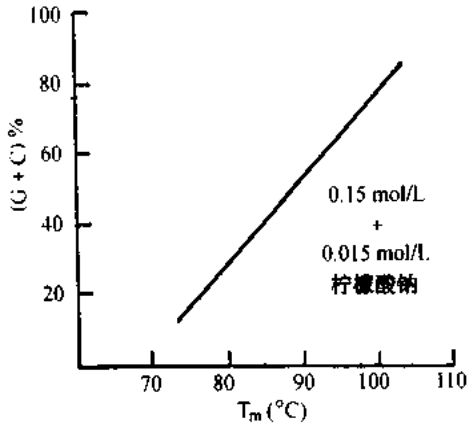


图 9-19 各种 DNA 分子的 Tm 值与 G+C 含量的关系。(引自 D. Freifelder: Molecular Biology, 1983, Fig4-7)

影响变性的因素有外部条件和内部结构二类。外部条件如温度和正离子的浓度。前面已介绍了正离子对磷酸骨架负电荷的中和及屏蔽以及对碱基疏水性的增加,所以影响到双链的打开。在正离子浓度低于 0.4mol/L,单价阳离子增高 10 倍,T_m 增加 16.6℃(图 9-19)。本身的结构主要是和 GC 含量及分子类型有关。当 GC 的含量上升 1%,则 T_m 上升 0.4℃,**马默多蒂**(Marmur-Doty)**关系式**:$T_m = 69.3 + 0.41(G+C)\%$,或 $GC\% = (T_m - 69.3) \times 2.44$。

通过这个关系式,我们可以通过 T_m 的测定,推出 GC 含量。或已知 GC 含量,可推出 T_m。在进行 PCR 引物设计时要求一对引物的 T_m 相近,就需考虑 GC 含量和引物的长度来调整其 T_m。另外 DNA 分子的类型不同 T_m 也不同,环状比线状稳定,所以其 T_m 高于线状分子的 T_m。

复性动力学的研究可以用来推测原核基因组的大小,以及区分单一顺序和复杂顺序等。对于复性反应我们可以简单地写成,2ssDNA $\underset{K}{\overset{K_2}{\rightleftharpoons}}$ dsDNA,即两条单链 DNA 形成双链 DNA,K 为反应的速率常数,基大小由溶液的阴离子浓度,反应温度及 DNA 片段长度的大小及复杂性所决定。此反应符合化学动力学二极反应速率方程:

$$\frac{dc}{dt} = -Kc^2$$

dc:浓度的变化;dt 是时间的变化;c 是 DNA 的浓度。此公式表示:反应速率和浓度平方成正比。

现将公式改写为

$$\frac{dt}{dc} = -\frac{c^{-2}}{K}$$

移项 $dt = -\frac{1}{K}c^{-2} \cdot dc$

积分 $\int dt = \int -\frac{1}{K}c^{-2} \cdot dc = -\frac{1}{K}\int c^{-2}dc$

$$t = \frac{1}{K}c^{-1} + 常数 \qquad (1)$$

当开始时时间 $t = 0$,初浓度为 C_0 则常数 $= -\frac{1}{K}C_0^{-1}$ (2)

将(2)代入(1)　$t=\frac{1}{K}C^{-1}-\frac{1}{K}C_0^{-1}=\frac{1}{KC}-\frac{1}{KC_0}$

移项　$\frac{1}{KC}=t+\frac{1}{KC_0}=\frac{KC_0t+1}{KC_0}$

移项得　$\frac{C}{C_0}=\frac{1}{1+KC_0t}$

初始时单链 DNA 的浓度 C 即是 C_0　$\therefore\frac{C}{C_0}=1$ 全部复性时 $\frac{C}{C_0}=0$

那么复性进行一半时 $\frac{C}{C_0}=\frac{1}{2}=\frac{1}{1+KC_0t_{1/2}}$

$\therefore C_0t_{1/2}=\frac{1}{K}$

因 K 是和 DNA 分子的大小呈反比的($K\propto 1/N$)所以复性反应进行一半时的浓度和时间的乘积和 DNA 分子大小成正比：

$C_0t_{1/2}\propto N$

根据复性动力学公式我们可以知道：

(1)单链 DNA 的浓度"C"随着时间"t"的增大而减小；

(2)反应速率取决于初始单链的浓度 C_0；

(3) 我们可以用反应浓度为纵坐标，以 $C_0t_{1/2}$的对数为横坐标绘制出各种 DNA 的

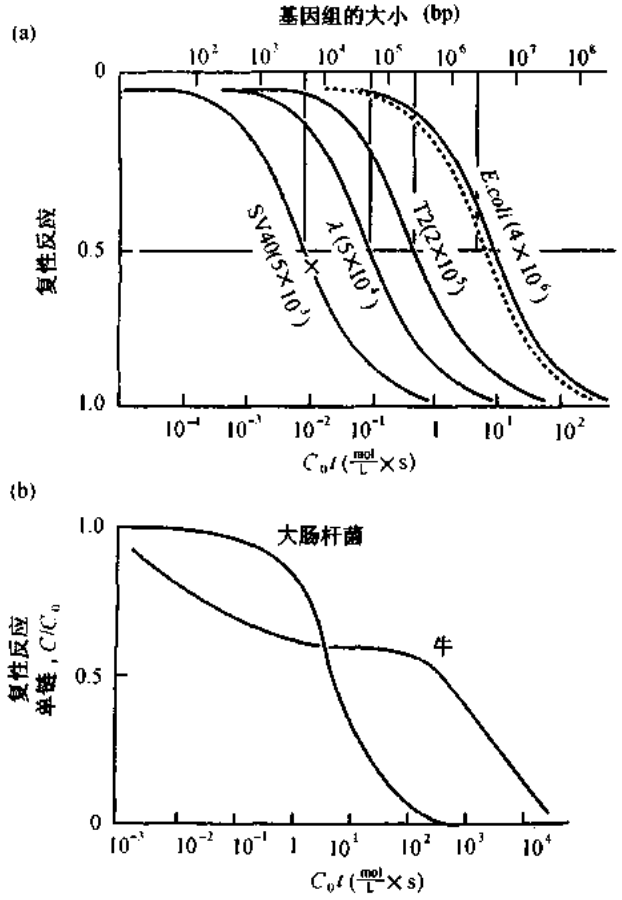


图 9－20　不同生物的 C_0t 曲线。(a)几种原核生物的 C_0t 曲线；(b)*E. coli* 和牛的 C_0t 曲线。(引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 10－25)

复性曲线；

(4) 因为 $C_0t_{1/2}$和基因组的大小成正比，我们就可以通过 $C_0t_{1/2}$值的测定来推测原核生物和噬菌体的分子量。已知大肠杆菌的基因组为 4.2×10^6 bp，$C_0t_{1/2}$ = 9M.sec，那么我们只要测出某原核生物的 $C_0t_{1/2}$，用公式：

$C_0t_{1/2}$(任何基因组 DNA)/9M.sec = 任何基因组大小/4.2×10^6 bp

就可以计算其基因组的大小。

(5)不同的原核生物因基因组大小不同，复性曲线也不同(图 9-20)。

(6) 可以用以区分真核生物和原核生物的基因组。真核生物的 DNA 有重复顺序，而原核生物多为单一顺序。单一顺序的复性曲线常只有一个拐点，而重复顺序常有多个拐点(图 9-21)。即使有些真核生物中的单一顺序只有一个拐点，但其 C_0t 比值变化

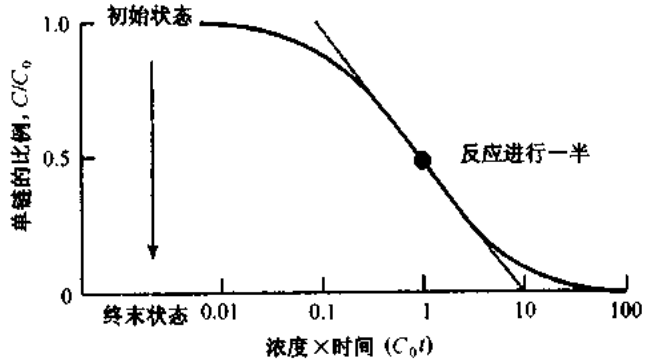


图 9-21 单一顺序复性曲线的判断。DNA 开始时是单链，变性终止时是双链。复性曲线的 80%跨越了 $21\log C_0t_0$(引自 Russell, P. J.: Genetics, 3rd ed. 1992, Fig 10-34)

范围和原核也不同。我们可以通过曲线的拐点(即反应进行一半时的这一点)做切线，在横坐标上的两个交点即为 C_0t 比值变化范围。原核生物的 C_0t 比值为 100 而真核生物的 DNA 的 C_0t 比值将大于 100，复性反应也受到以下因素的影响：①DNA 片段的大小；②DNA 的浓度；③DNA 复杂性，即原核为单一顺序，所以 DNA 复性较为稳定，而真核 DNA 具有各种重复顺序，所以较为复杂。④温度。复性温度一般比 T_m 低 25℃为最佳。(5)盐的浓度。复性时要求盐的浓度达到足够高，如 NaCl 在 0.15~0.50mol/L。

在复性反应中如何知道单链已经结合成双链了呢?一般我们可以通过以下两种方法来检测：

(1)减色效应(hpochromic effect)

和增色效应相反，由于双链逐渐增多，对 260nm 的紫外光吸收会相应减弱。因此测定光密度，即 OD 值(optical density)即可。DNA 从单链变成双链，OD_{260}减少 30%。

(2)羟基磷灰石柱层析

羟基磷灰石的特点是对双链 DNA 吸附较牢，不易吸附单链，我们将样品过柱，然后用不同浓度的缓冲溶液洗脱，就可分别收集单链和双链 DNA，计算相对浓度。

第五节 分子杂交

分子杂交(molecular hybridization)是用一个 DNA 单链或 RNA 单链与另一被测 DNA 单链形成双链，以测定某特异顺序是否存在。这种方法已成为遗传学和分子生物学等

生命学科中最为普遍和重要的方法之一。分子杂交的方法多种多样，但其共同特点是：①都是应用复性动力学原理；②都必须有探针(probe)的存在。所谓探针就是用同位素或非同位素如荧光染料，生物素等标记的短片段特异 DNA 或 RNA 顺序。我们常用的分子杂交方法有以下几种：

一、原位分子杂交(in situ hybridization)

原位分子杂交有两种，一种是玻片原位杂交，另一种是膜上原位杂交。玻片原位杂交一般都先要制片，如中期染色体片和组织切片等，片子制好后不染色，放在缓冲溶液

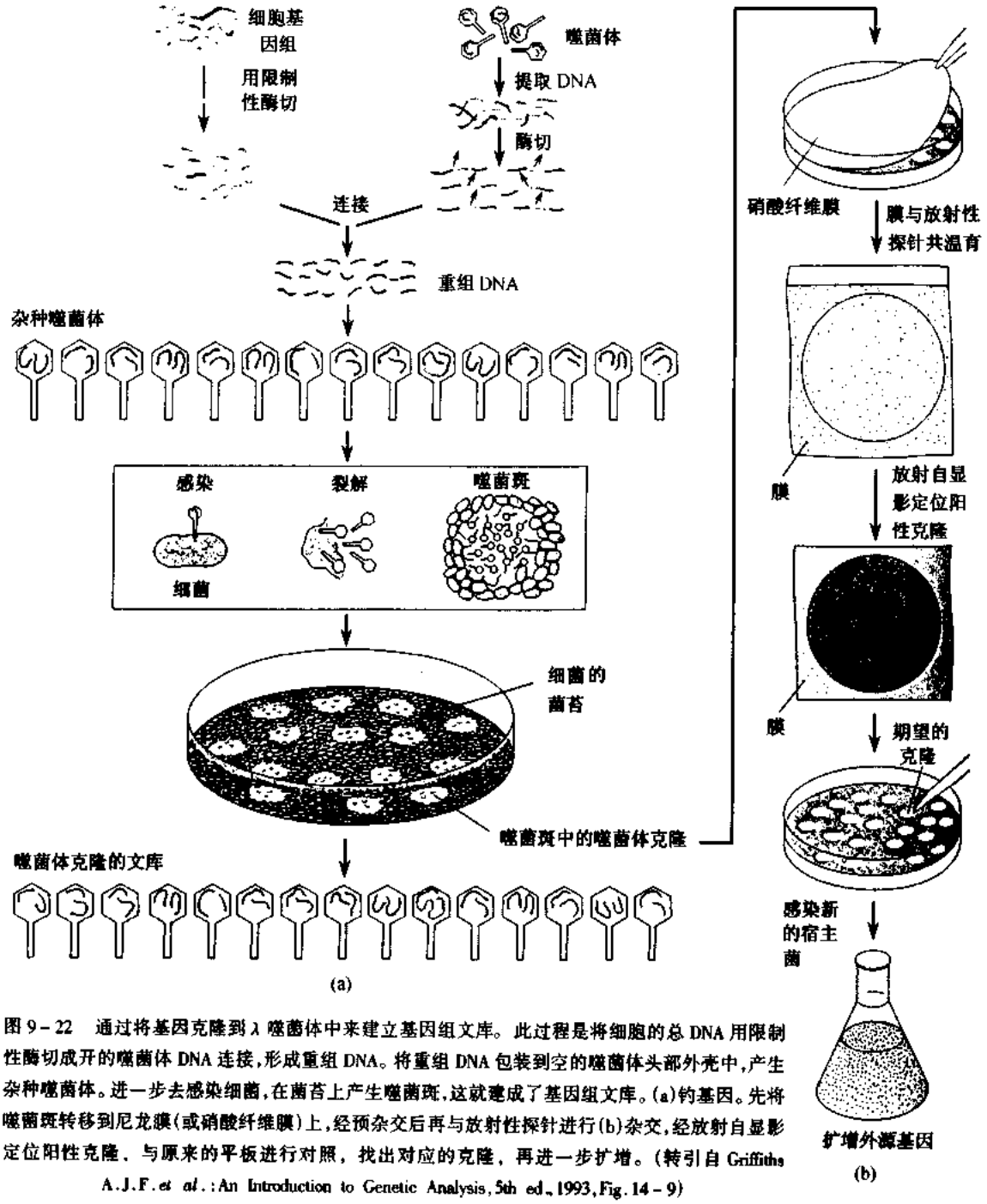


图 9－22　通过将基因克隆到 λ 噬菌体中来建立基因组文库。此过程是将细胞的总 DNA 用限制性酶切成开的噬菌体 DNA 连接，形成重组 DNA。将重组 DNA 包装到空的噬菌体头部外壳中，产生杂种噬菌体。进一步去感染细菌，在菌苔上产生噬菌斑，这就建成了基因组文库。(a)钓基因。先将噬菌斑转移到尼龙膜(或硝酸纤维膜)上，经预杂交后再与放射性探针进行(b)杂交，经放射自显影定位阳性克隆，与原来的平板进行对照，找出对应的克隆，再进一步扩增。(转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 14－9)

中缓缓变性，然后再加入探针让其复性，将没有结合的探针充分洗脱，若是同位素标记探针，就须在片子涂一层乳胶或覆盖上一张同样大小 X 光片进行放射自显影，然后观察同位素的暴光点在组织细胞或染色体上的相对位置。如用荧光标记可用荧光显微镜直接观察。近年来已发展了多色荧光技术，同时可用不同的探针来进行杂交。这种方法可用来进行染色体的基因定位，或观察组织中不同的 RNA 的分布。果蝇早期的发育中，不同基因的控制就采用了这种方法进行研究。

膜上分子杂交是将菌落或噬菌斑影印在尼龙膜上（以前用硝酸纤维膜），这种膜只吸附单连 DNA。将影印好的膜用 NaOH 处理，这样不仅可以杀死细菌，同时可使 DNA 变性吸附在膜上，然后用无关的单链 DNA，如常用小牛胸腺 DNA 和鲑鱼精子 DNA 吸附到膜上的空白处，称**预杂交**，防止探针被吸附在背景上，干扰实验结果。膜经中和后再将同位素探针和膜放在缓冲溶液中缓缓复性，经放射自显影来确定阳性菌落。这种方法常用来从基因文库（gene bank 或 gene library）中来钓基因（图 9－22）。

二、斑点杂交

斑点杂交（dot blotting）也叫狭缝杂交。是由 Southern blot 衍生而来。方法是将某种生物的总 DNA 直接点滴在尼龙膜上，或者通过一个小的长形狭缝印在尼龙膜上，然后将膜变性处理，预杂交后，经中和、洗脱、干燥。再将其和探针放入复性缓冲溶液中缓缓复性，洗涤干燥后进行放射自显影，观察结果。这种方法是用来初步探测某种生物中含有一种特殊的基因或顺序来分析 DNA 样品之间的同源性。此法因有假阳性的存在，所以发现阳性斑后还要进一步做 Southern blot 加以验证。

三、Sou thern 杂交

萨瑟杂交是 E.M.Southern 于 1975 年首先建立的，故称为 Southern 杂交，该方法经萨瑟印迹法将胶上的电泳条带吸印尼龙膜上，然后和 DNA 探针杂交（图 9－23），用这种方法可以制作染色体的**物理图谱，限制性片段长度的多态性，及动物园印迹**（Zoo blot）（用一个物种的 DNA 探针和别种动物的 DNA 进行 Southern blotting）等。

四、Northern 杂交

Northern 杂交的技术和 Southern 杂交相似，作者并不姓“Northern”，因“Southern”意为“南方”，故戏称自己的分析方法为“北方”（Northern）杂交。Northern 杂交与 Southern 杂交主要的不同之处在于是用来检验 RNA，而不是 DNA。首先是提取某种生物或组织的总 RNA 或 mRNA，然后用含有变性剂的琼脂糖凝胶电泳分离 RNA，变性剂的作用是防止 RNA 自我退火形成局部双链，影响泳动率，干扰实验结果。电泳分离后再将凝胶上的 RNA 带印迹到尼龙膜上，在液相中和标记的核酸探针进行杂交。用此方法可以测定某基因表达的时空特异性。

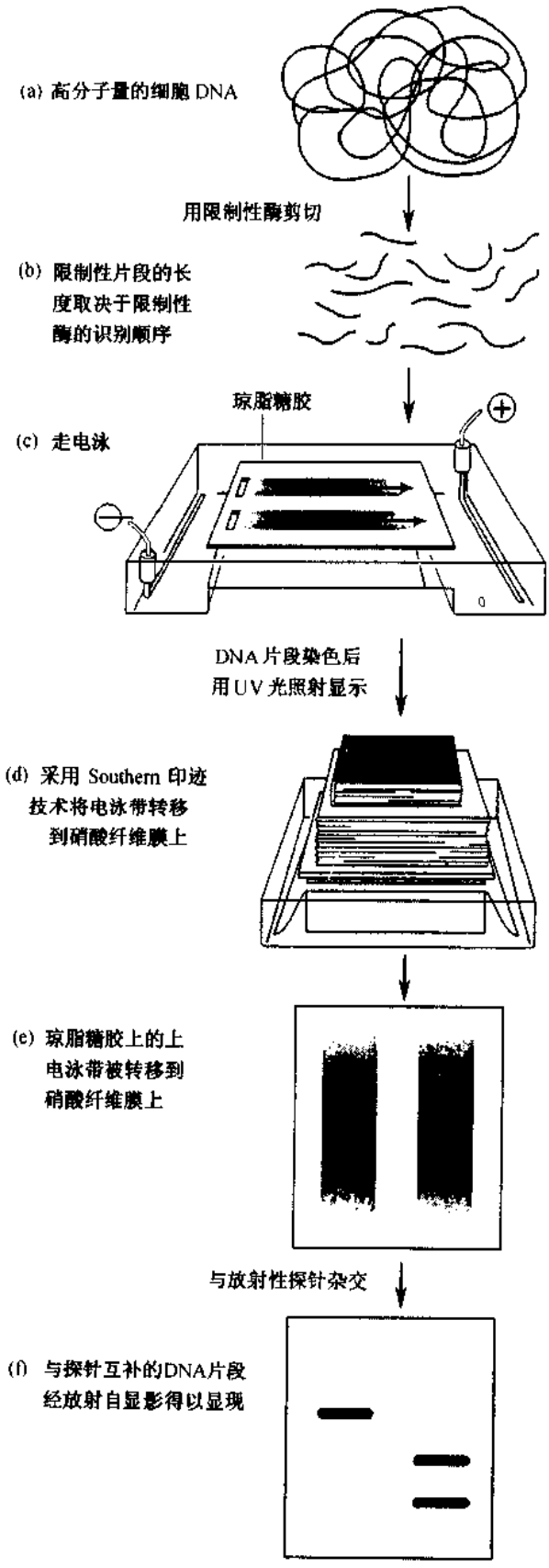


图 9 - 23　Southern 杂交实验。(引自 Russell, P. J: Genetics, 3rd ed., 1992, Fig. 15 - 18)

五、Western 印迹

既然有了“南方”杂交，“北方”杂交，那么随之而来的就是“西方”杂交（Western blotting），但 Western 印迹和前面的几种杂交都不相同，它是用来测蛋白而不是检测核酸，在此一并简单介绍一下，此法是由电场的作用下将聚丙烯酰胺的电泳胶上的蛋白质带转移到尼龙膜上，以亲和反应或免疫反应或结合反应测定蛋白质的技术。

六、电镜观察

先进行分子杂交，再通过电镜观察可以进行基因定位，内含子的探测以及缺失的确定等。断裂基因就是用这种方法发现的（见第 10 章）。电镜观察有两种，一是**异源双链定位法**（heteroduplex mapping），二是**R－环定位法**（R-loop mapping）。异源双链定位法是将异源 DNA 双链变性后进行分子杂交，然后制片在电镜下观察，可发现环形不配对的部分，表明可能存在缺失（见第十章）（图 9－24a）。R－环法是将标记的 mRNA 和变性后的待测双链 DNA 进行杂交，制片后，可以观察到 R－环（图 9-24），这是由于 RNA 形成的双链比原来的 DNA 双链更为稳定，将一条 DNA 单链置换了出来，表明该 mRNA 基因就位于 R－环这个位置。取名为“R－环”有两层意思：①是取代、置换之意，用英文单词“replacement”的第一个字母来表示；②显示了取代环的形状（图 9－24b）。

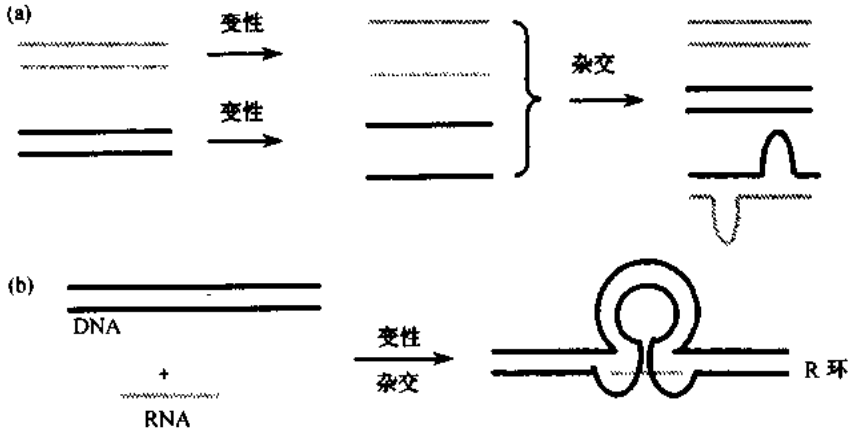


图 9－24　异源双链定位法(a)和 R 环作图(b)。

第十章

基因组和染色体

基本概念

基因组(genome)是指一个细胞中遗传物质的总量。真核细胞常为**二倍体**(diploid),所以其基因组是指细胞中的单套染色体及其上的基因。

在原核生物和病毒中基本遗传物质是 DNA,其染色体含有大量的和其结合的蛋白质。

细菌的染色体是环状双链的 DNA 分子,呈超螺旋状态。在 *E.coli* 的染色体上已鉴别出 100 个独立成环的超螺旋 DNA 区域。

真核细胞中全套染色体称为**核型**(karyotype)。核型具有种的特异性。

在原核染色体中,DNA 的量或真核细胞中单倍体 DNA 的量,称为 C 值(C value)。C 值和有机体的结构或组成的复杂性并没有直接的关系。

真核细胞的核染色体是由 DNA、组蛋白和非组蛋白组成的。每条染色体都是由染色质丝结合在染色体骨架上所形成。组蛋白在有机体中从一个细胞到另一细胞都是不变的,而非组蛋白在不同类型的细胞中明显不同。

真核染色体上存在大量的 DNA,它和组蛋白构成**核小体** (nucleosome),再高度盘绕成**染色质丝** (chromatin fibre),直径为 30nm。每条染色体上含有大量侧环,这是由染色质丝附着在蛋白质骨架上形成的环状区域。

每个真核染色体的**着丝粒**区是在有丝分裂和减数分裂中负责复制后的染色体精确分离进入子细胞。着丝粒的 DNA 序列具有种的特异性。

端粒(telomeres)在染色体的末端,含有短而简单的**衔接重复序列**(tandomly repeated sequences)。也具有种的特异性。

原核基因组含有大量**单一序列**(unique - sequences),DNA 仅有少量的重复顺序和基因。真核生物的基因组含有单一序列和重复序列。真核生物的重复序列存在复杂谱系是不难想象的。

原核生物的细胞中除了主染色体以外,还含有各种质粒和转座因子。质粒常为双链环状 DNA,可独立复制,有的既可以游离于细胞质中,也可以整合到染色体上。转座因子一般都是整合在基因组中。

真核生物除了核染色体以外,还存在细胞器 DNA,如线粒体和叶绿体的 DNA,为双链环状,可自主复制。有的真核细胞中也存在质粒,如酵母和植物。

第一节　原核生物基因组的结构特点

原核生物的染色体分子量较小,以 *E.coli* 为例,约为 2.4×10^9 Da,长度为 4200kb 左右(1300μm),形态为闭合环状,约编码 2000 个基因,原核生物的 DNA 位于细胞的中央,称为**类核**(nucleoid)(图 10 - 1)。*E.coli* 的类核由**支架**(scafford)和向四周伸出的 100 个 DNA 环组成,支架的形状因染色体而异,长度为 3 ~ 5μm,支架是含 RNA 和蛋白质的复合结构,四周的每个环就是一个独立的功能区,长 40kb,13μm(图 10 - 2)。*E.coli* 在体内(*in vivo*)其基因组内含负超螺旋,每 200bp 就有一个负超螺旋($\sigma = 0.05$),即基因组中含 5%的负超螺旋。每个功能区的末端保持超螺旋状态,而且一个区的超螺旋

不影响另一个区的超螺旋，功能区的相对独立性使得同在一个环状染色体上的基因可以独立表达和调控。基因组内的超螺旋以两种状态存在：①自由状态的超螺旋不受束缚，可在环内传递张力；②蛋白结合在 DNA 上超螺旋受到束缚，不能传递张

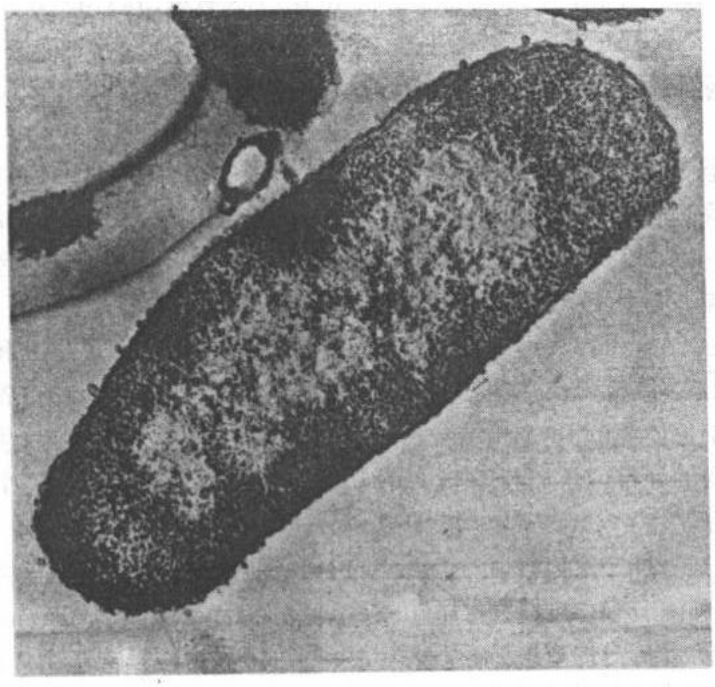


图 10－1　*E. coli* 的类核。照片可见到细胞内有一团紧密的组织，此就是类核。（转引自 B. Lewin: Genes, Ⅵ 1997, Fig. 26－3）

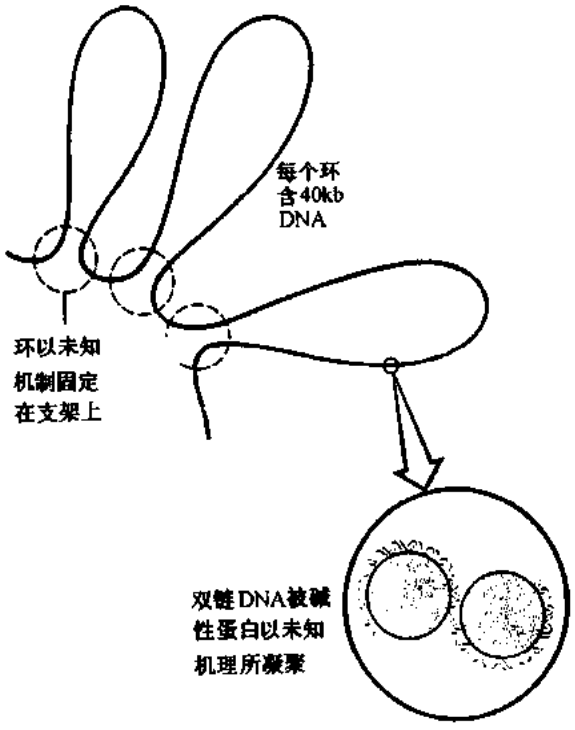


图 10－2　类核的结构。（仿 B. Lewin: Genes, Ⅵ 1997, Fig. 26－5）

力。

细菌的染色体约有 75% 的 DNA 编码基因，余下 25% 的 DNA 为基因间的 DNA (intergenic DNA)。有一部分基因间的 DNA 是有重要功能的，如复制原点 (origin of replication)，另一些基因间区域可能涉及到和 DNA 包装蛋白的相互作用。

一、类核的结构

原核生物的 DNA 虽不像真核生物那样和组蛋白结合形成四级结构构成染色体，但也并非是完全裸露的，结构蛋白怎样和细菌的染色体结合呢？当 *E.coli* 的细胞被裂解后，类核区 DNA 就释放出去(图 10－3)，形成环状，但并未发现这些环状的双链 DNA 延伸的游离端，这是因为 DNA 和蛋白质结合变得更为紧密之故。

各种 DNA 结合蛋白表面上和真核生物的染色体蛋白相似，那么我们用什么标准来确定一个 DNA 结合蛋白在类核中是否起构成染色体的作用呢？首先它应有足够的数量和整个染色体组的 DNA 相结合；第二，当编码这些蛋白的基因发生突变时将会导致染色体结构的异常或蛋白与 DNA 结合功能的消失。目前还没有一个候选的蛋白能满足这些条件。

Hu 蛋白是一种二聚体，相当于真核的 H_2B，能使 DNA 压缩 、盘绕成念珠状的结构，它还可以刺激 DNA 的复制(见第 10 章)。它和**整合宿主因子**(integration host factor, IHF)相关。**IHF** 也是一种二聚体，和一些特殊的重组反应包括 λ 噬菌体的整合、切离有关，有助于 λ 噬菌体和宿主染色体在 *att* 位点的配对、重组。在形成蛋白复合体中，它起到构建的作用。

表 10－1　*E.coli* 含有的各种 DNA 结合蛋白

蛋白	结构	功能	含量/每细胞	相当于真核蛋白	基因
HU	α和β亚基，每个9kDa	使DNA压缩、类核凝聚，刺激复制，和IHF有关	4万个二聚体	H_2B	*hupA.B*
H	两个相同亚基，各28kDa	促使双链的互补、复性	3万个二聚体	H_2A	?
IHF	α：10.5kDa；β：9.5kDa	有助于*att*位点配对重组	?	?	*himA.D.*
H1 (H－NS)	15kDa亚基	和DNA结合，与DNA拓扑结构有关	1万	?	*osmZ bglY*
HLP1	17kDa单体	?	2万	?	*fisA*
P	3kDa亚基	?	?	鱼精蛋白(DNA 结合蛋白)	?

H 蛋白由两种相同的亚基组成，其氨基酸组成与真核生物组蛋白的 H_2A 相似，有

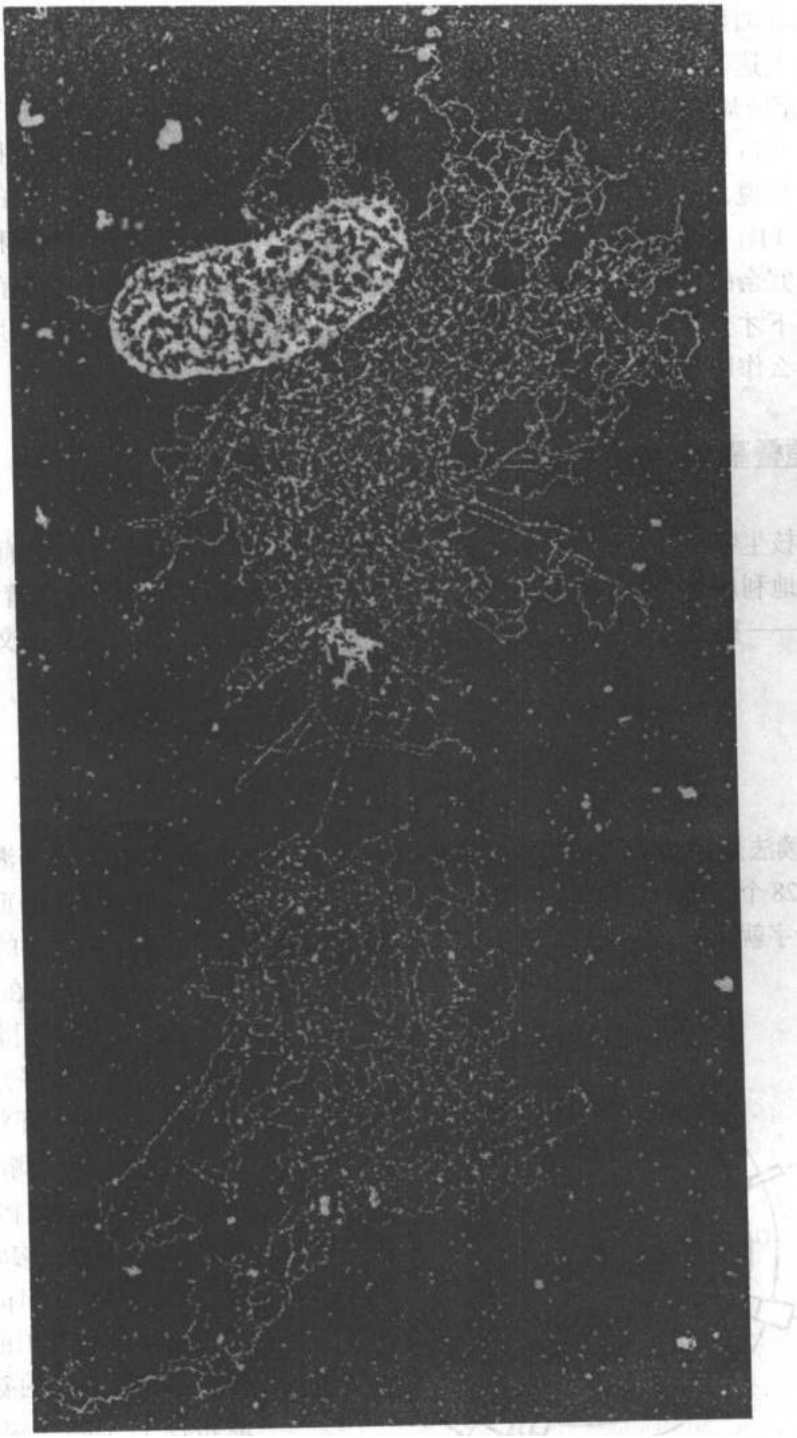


图 10－3　*E. coli* 细胞裂解时类核溢出胞外形成环状纤丝。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig. 26－3)

促进互补的 DNA 单链结合成双链的作用。

编码 Hu 两个亚基的基因 *hupA*, *B* 发生无效突变时，影响不大，仅失去了冷敏感表型和 DNA 的某些超螺旋。结果表明 Hu 在类核的凝聚中起一定的作用，但不像 IHF 在 λ 噬菌体重组时可起到的那种帮助特异顺序配对的作用。

H1 蛋白也称 H-NS 蛋白，可以和 DNA 结合，而优先和弯曲的序列相互作用。其基因(*osmZ*, *bglY*, *pilG*)若发生突变就会引起外表的变化，已鉴定出这些基因是不同

系统的表面调节基因。此表明 H1 蛋白可能和 DNA 的拓扑学结构有关，从而影响到基因的表达。

P 蛋白基因的编码的产物类似于真核的鱼精蛋白(protamines)，鱼精蛋白是可以和某些精子的 DNA 结合，从其序列来看它是一种 DNA 结合蛋白，但其功能和数量都不知道。

按理说，若缺乏类核结构所需的结合蛋白，将会严重地影响微生物的生命力，但删除 Hu 和 H1 的基因时，对细菌的影响却不大，这是为什么呢？对此有两种解释：①这些蛋白是冗余的，它们其中的任何的一个都可以作为另一个的亚基，只有在全部被删除的情况下才会对类核结构产生严重的影响；②我们尚未搞清这些蛋白对类核的组成究竟起什么作用。

二、重叠基因

原核生物和一些病毒或噬菌体的基因组比较小，核苷酸对是极其有限的，那么怎样有效地利用这些有限碱基来编码更多的遗传信息呢？在生物中存在着一种十分巧妙的机制——**重叠基因** (overlapping gene)。它的道理就像我们古代的回文诗一样，如：

其读法是：采莲人在绿杨津，在绿杨津一阙新，一阙新歌声激玉，歌声激玉采莲人。本来需要 28 个字才能表达完的一首诗，现在利用前后两句之间的几个字的重叠，结果只用了 14 个字就完成了。重叠基因也正是如此，利用碱基的重叠来编码更多的信息。

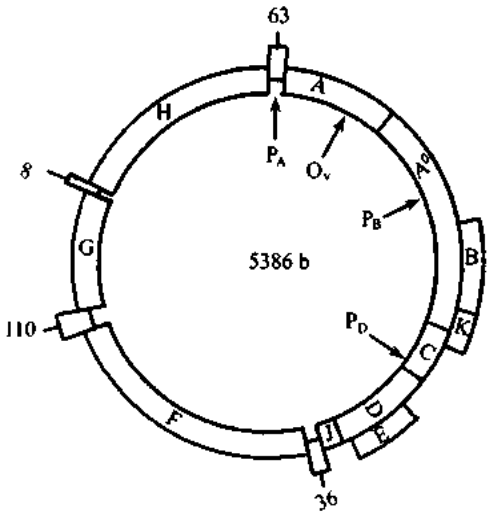


图 10－4　ΦX175 的环状基因组。*B* 基因在 *A*· 基因内，*K* 基因跨在 *A*· 基因和 *C* 基因的边界上。*E* 基因在 *D* 基因内，显示基因重叠。

重叠基因是在 1977 年首先发现的，当时英国著名的科学家 Sanger 建立了测序方法，他就用这种测序方法对环状单链的噬菌体 *ΦX*174 进行了测序。结果测出其基因组由 5386 个核苷酸组成，共有 11 个基因，构成 3 个转录单位，由 3 个启动子(pA，pB，pD)启动(图 10－4)。基因的产物都已被分离，它们所含的氨基酸已远远地超过了 5386 个碱基所能编码的量，即 *ΦX*174 含有的 5386nt 最多能编码 1795 个氨基酸，若每个氨基酸的平均分子量为 110，则总的蛋白质分子量为 197 000Da，但实际测出的蛋白质总分子量却为 262 000Da。将全部 DNA 序列和蛋白质的氨基酸顺序进行比较，发现 B 基因在 A^0 基因之中，*K* 基因跨在 *A* － *C* 两基因的连接处，和 A^0 基因的尾部，*C* 基因的首部相重叠。*E* 基因在 *D* 基因内部。类似的情况在别的噬菌体如 G4(图 10－5)、微小病毒和 SV40 中也有发现。重叠基因不仅可经济利

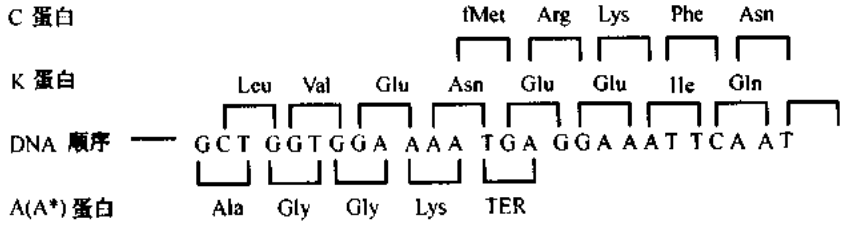


图 10-5 噬菌体 G4 基因组中 *C*, *K* 和 *A*(*A**)基因的重叠。(引自 G.N.Godson, B.G.Barrell *et al*.: 1978, Nature, 276: 236)

用基因组，而且可能起表达调控的作用(见第 16 章)。

重叠基因仅在噬菌体和病毒中存在，在真核生物中尚未发现重叠基因。这可能因为前者基因组比较小，但又必须要编码一些维持其生命和繁殖的基因，在选择的压力下，保留了这种重叠基因的形式。

三、质粒

几乎所有的细菌中，除了染色体之外，细胞中还存在一些环状的 DNA 分子，叫做**质粒**(plasmid)。质粒上常带有一些特殊的基因，如抗生素的抗性基因。质粒具有自己的复制起始区，能自主复制。有些小质粒是使用细胞的酶来复制，而大型质粒本身就带有编码复制酶的基因，这样就可以依靠自己的复制酶来复制。另外一些质粒它们可以整合到宿主的染色体中，和细菌染色体同步复制。

在细菌中有很多不同的质粒，不同的细菌含有不同的质粒。质粒根据它们所带有的基因以及赋于宿主细胞的特点可以分为五种不同的类型：

抗性质粒 (resistance (R) plasmids)　它们带有抗性基因，可使宿主菌对某些抗生素产生抗性，如对氨基苄青霉素，氯霉素等产生抗性。不同的细菌中也可含有相同的抗性质粒，如 RP4 质粒在假单胞菌属和其他细菌中都存在。R 质粒还可以通过感染的形式在不同种的细菌中传播。

致育因子 (fertility (F) plasmids)　可以通过接合在供体和受体间传递遗传物质。F 因子约有 1/3 的 DNA 构成一个转移 DNA 的操纵子，约 35 个基因，负责合成和装配性伞毛。这就是 DNA 转移区域，受 *traJ* 基因产物的正调节。它还具有重组区和复制区。重组区含有多个插入顺序，通过这些插入顺序进行同源重组。在复制区有两个复制起始点：一个是 *oriV*，供给 F 因子在宿主中自主复制时使用；另一个是 *oriT*，供接合时进行滚环复制的起始点。

Col 质粒　带有编码大肠杆菌素 (colicins) 的基因。大肠杆菌素可杀死其他细菌。如 *E.coli* 中的 ColE1。

降解质粒(degradative plasmids)　这种质粒编码一种特殊蛋白，可使宿主菌代谢特殊的分子，如甲苯或水杨酸。

侵入性质粒(virulence plasmids)　这些质粒使宿主菌具有致病的能力。如 Ti 质粒，此是在根癌农杆菌 (*Agrobacterium tumefaciens*) 中发现的，现经过加工用来作植物转基因的一种常用载体。

质粒的大小很不相同，最小的长约 1kb，而最大的可达 250kb。质粒在细胞中的拷

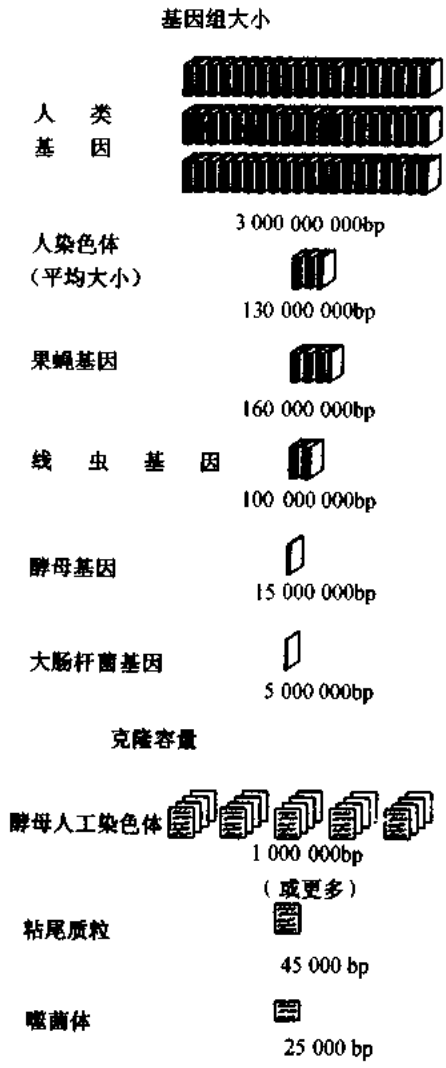


图 10－6　人类和基他生物基因组大小的比较。上：几种 DNA 标本的近似大小，其 DNA 序列的含量以书本表示，每本书约相当于 5 千万碱基对（bp）的 DNA 序列。下：酵母人工染色体（YAC）、粘尾质粒及噬菌体的克隆容量。其 DNA 序列的含量 m 书页表示，每页约相当于 5 万 bp 的 DNA 序列，约为普通杂志每页字数的 6 倍。（引自（Eric D.Green 等：《人类基因组工程临床医学的前景和意义》，美国医学会杂志中文版 1992,11(3),P144）

贝数也不尽相同。有的拷贝数很少，只有一，二个拷贝，称为**严紧型**（stringent）**质粒**，另一些具有较高的拷贝数，约有 10 个以上，称为**松弛型**（relaxed）**质粒**。有些质粒不能共存于一个细菌的细胞内，这种特征称为**质粒的不相容性**（plasmid incompatibility）。共存于同一种细菌中的质粒必须属于不同的不相容群。

在细菌的基因组中还存在着一些转座因子，它们有着特殊的结构，并可以在基因组中移动（见第 23 章）。

第二节　真核基因组的复杂性

真核生物的基因组分子量很大，约为 3×10^{2} Da，长度为 2×10^{6} kb，形态为线状，约编码 100 万个以上的基因，染色体和组蛋白结合形成四级结构。人类的每个细胞在二倍体的核中 DNA 长约 30 亿个碱基对，估计编码 10 万个基因，每条染色体含碱基 8 千万至 3 亿不等。现已有 5 千个基因被编目，1900 个基因已进行了染色体定位，600 个已被克隆分离出来（图 10－6）。若将一个细胞中每条染色体的 DNA 首尾彼此相接，全长约 200cm。Renato Dulbecco 于 1986 年首先提出了“**人类基因组计划**”，这一计划约 10～15 年完成，耗资 30 亿美元，其宏伟的程度堪与 Manhatto 原子弹计划和 Appolo 登月计划相提并论。经多年的激烈争论，终于在 1990 年开始实施，由美国国立卫生研究院（NIH）和能源部的两个人类基因组国家研究中心负责协调。要将人类基因组的全顺序都测定出来。这一工作进展很快，1999 年国际人类基因计划联合小组完整地破译出人体第 22 号染色体的遗传密码。2000 年 5 月国际人类基因计划完成时间再度提前，预计从原定的 2003 年 6 月提前到 2001 年 6 月。2000 年 5 月国际科研小组宣布，他们已基本完成了人体第 21 号染色体的测序工作。2000 年 6 月 26 日美、英、日、德、法、中六国政府和有关科学家共同宣布人类基因组工作草图绘制成功。Watson 指出“现在我们有了人类生命的说明书”。这一工作结束之日，将使人们深刻地了解人类生老病死的奥秘，大大增强人们对遗传病、肿瘤、心血管病，以及最新发现的一

些病毒感染的疾病的治疗和预防的能力，对其他的真核生物也会起到触类旁通的作用。但也可能会引起种族歧视等负面效应。

将人类的染色体完整地分离出来进行研究是十分困难的。在分离纯化时常引起断裂。分离真核的染色体是在有丝分裂的中期，分离 DNA 是在间期的核中提取。从细胞中获得的 DNA 和蛋白质的复合物称为**染色质**(chromatin)。在各种真核细胞中鉴定染色质是容易的。

染色质有两种类型：**常染色质**(euchromatin)和**异染色质**（heterochromatin)。常染色质在细胞分裂时凝聚，而在间期时不盘绕。异染色质在细胞周期皆保持凝聚状态存在。此通过其染色的性能可得到证明。常染色质一般是有活性的，而异染色质一般都是没有活性的。这是由于它不含有基因(组成型异染色质，如着丝点区域)或抑制了所含基因的表达(兼性异染色质，如失活的 X 染色体)。异染色质在 S 期复制时比常染色质要晚一些，此是由于异染色质中高度浓缩的结果。

异染色质在各种真核生物中都发现于着丝点附近及端粒上，而在其他位置时常具有种的特异性。两种异染色质也是可以区分的。**组成型异染色质**（constitutive heterochromatin)总是没有活性的，且位于染色体对的同源位点上。着丝点和端粒皆属于此。**兼性异染色质**(facultative heterochromatin)是可以变成异染性的染色质。它可能含有基因，但染色质凝聚时这些基因也随之失活，女性中有一个 X 染色体失活而成为巴尔小体（Barr bodies）就是一例。

在染色质中和 DNA 结合的两类蛋白：**组蛋白和非组蛋白**，他们在确定染色体的物理结构中起了重要的作用。正如我们将要看到的 DNA 总是包在组蛋白分子的四周，而非组蛋白再和这个复合体结合。各种研究表明有的非组蛋白在染色体中起到结构的作用。如果我们将染色体中的组蛋白除去的话，那么 DNA 就解开且从复合体中释放出来，但非组蛋白骨架以染色体的形状留了下来(图 10－7)。

一、组蛋白

组蛋白是和染色体相联的最主要的蛋白质。它的作用是和染色体中的 DNA 的负电荷结合。组蛋白是比较小的碱性蛋白；在细胞正常 pH 值时，组蛋白带有正电荷，这样它们就可以和 DNA 结合，这个正电荷主要存在于碱性氨基酸 Lys 和 Arg 的 $-NH_3^+$ 上。其实，组蛋白约含 25% 的 Arg 和 Lys。比其他蛋白的 Arg，Lys 的含量都多。

和真核 DNA 结合的有 5 种类型的组蛋白：**H1，H2A，H2B，H3，H4**。组蛋白的总重量和 DNA 的总重量大致相等，这个总量和比例在各种真核细胞中都是如此。来自小牛胸腺 DNA 的组蛋白的特点如表(12－2)所示。此四个组蛋白的顺序 H2A，H2B，H3 和 H4 是通过各种生物测定所定。比较这个序列表明，在亲源关系较远的物种中这个序列仍十分保守，例如牛和蚕豆的 H4 组氨酸仅有 2 个氨基酸的差异[然而在酵母和四膜虫(*Tetrahymena*) 中 H3 组蛋白都有 11 个氨基酸的差异]。其实这四种组蛋白的保守性在已知的所有蛋白中算是最高的。氨基酸序列的保守性明显表明组蛋白在各种真核生物染色体的组成中起到相同的基本作用。

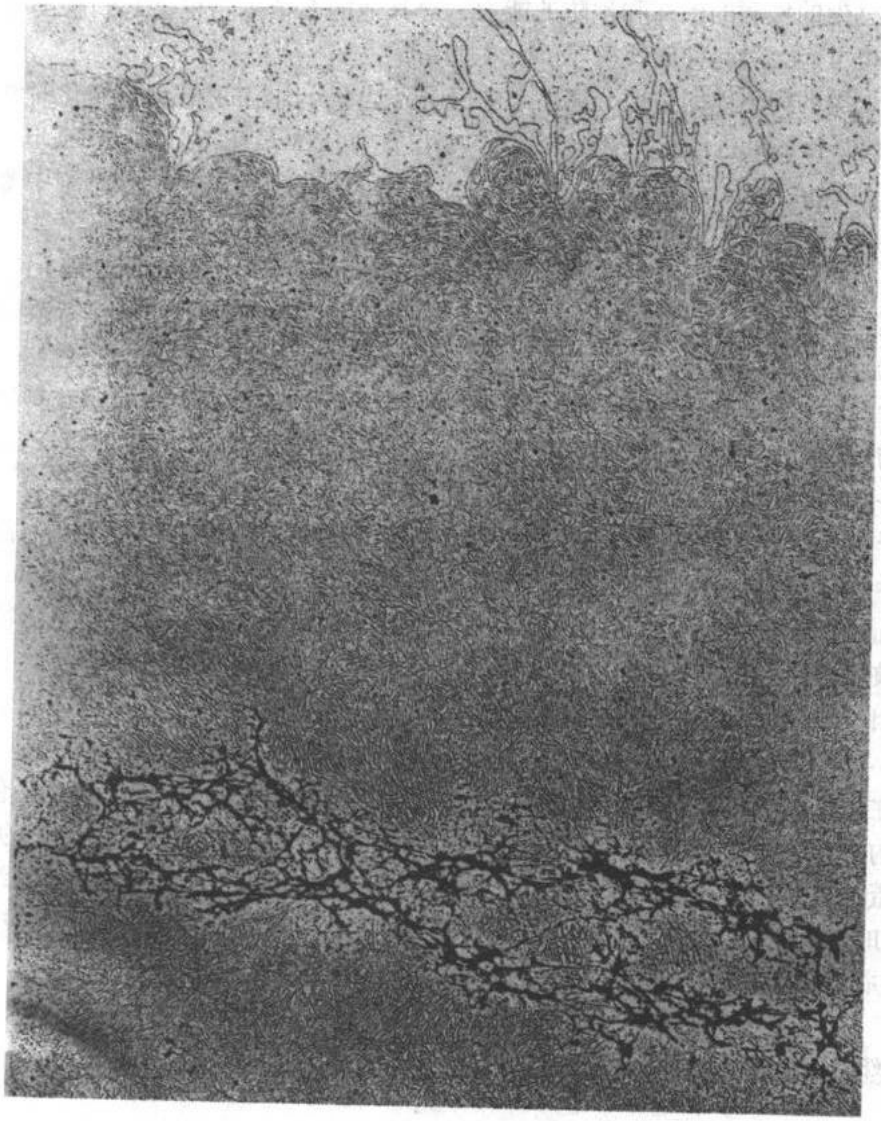


图 10－7　染色体非组蛋白的骨架。(转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993. Fig. 13－2)

表 10－2　小牛胸腺 DNA 的组蛋白的特点

组蛋白类型	碱性氨基酸		氨基酸数	分子量(Da)
	Lys	Arg		
H1	29%	1%	215	23,000
H2A	11%	9%	129	13,960
H2B	16%	6%	125	13,775
H3	10%	13%	135	15,340
H4	11%	14%	102	11,280

摘自 S.C.R. Elginand H. Weintraub, 1976, Annu. Rev. Biochem. 44: 725－726

各种细胞中的 H1 的组蛋白虽不完全相同，但紧密相关。这种组蛋白有一定的氨基酸核心区域。此是通过进化而保留下来的，但其余分子都缺少这种保守性。然而某些组织中 H1 并不存在，而被其余的组蛋白所取代。例如在鸟类的红细胞(和哺乳动物的红细胞不同，鸟类的红细胞有核)中 H5 取代了 H1。

二、核小体(nucliesome)

从 DNA 到染色体不论是形态还是长度都相差很大。人类最长的第一个染色体全长仅 10μm,但其 DNA 却长达 7.2cm;一个细胞核直径仅 5μm,在这样一个小小的空间中却要纳下全长近 200cm 的 DNA,人们不禁要问 DNA 如何形成染色体,纳入小小的核中。解决这个问题同样是由很多科学家差不多经过 20 年的努力,最终提出了为大多数能接受的模型——**侧环模型**。

早在 1956 年为双螺旋模型提供 X 光衍射证据的 Wilkins 和另一位科学家 Vittorio Luzzati 对染色质进行了 X 光衍射研究,发现染色质中具有间隔为 100Å 的重复性结构。蛋白质和 DNA 本身的结构从来不会表现出这种重复性。推测可能是组蛋白和 DNA 的结合方式迫使 DNA 折叠或缠绕成具有 100Å 周期的重复结构。

Clark 和 Felsenfeld 于 1971 年首先用葡萄球菌核酸酶(Staphylococcal nuclease)来作用染色质,发现有一些区域对核酸酶敏感,有一些则不敏感,不敏感的区域比较均一,这暗示染色体中存在着某些亚单位。接着 Hewish 和 Burgoyun(1973 年)用内源核酸酶消化细胞核,再从核中分离出 DNA,结果发现一系列 DNA 片段,它们相当于长约 200bp 的一种基本单位的多聚体。表明组蛋白结合在 DNA,以一种有规律的方式分布,以致产生对核酸酶敏感的只是某些限定区域。M.Noll (1974年) 用外源核酸酶处理染色质,然后进行电泳,证实了以上结果,他测得前三个片段的长度分别为205、405、605bp 长(图 10-8),每个片段相差 200bp,即染色质可能以 200bp 为一个单位。这正好和以下电镜观察的结果相映证。

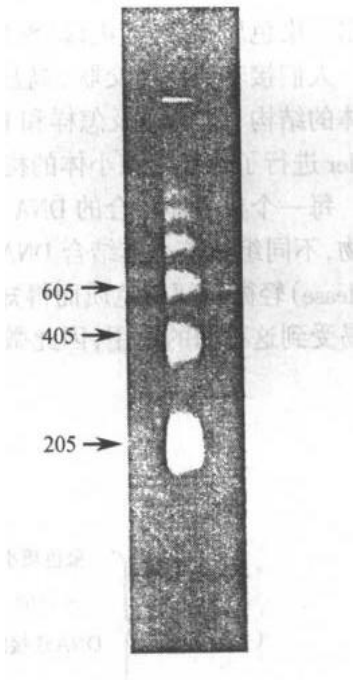


图 10-8 用微球菌核酸酶降解染色质后的电泳图谱。(转引自 B.Lewin: Genes, Ⅵ 1997, Fig.27-7)

与此同时 Olins 夫妇(1974)和 Pierre Chambon 等(1975)在电镜下观察到大鼠胸腺和鸡肝染色质的"绳珠"状结构(图 10-9),小球的直径为 100Å,Olins 并把这种小球称为 ν 小体(ν-body 即 nu body),有时译成钮体。

X 衍射图表明组蛋白的多聚体都是紧密相联,并无可容纳像 DNA 分子那样大小的孔洞,所以不可能由 DNA 之"绳"穿过组蛋白之"珠",而只可能是 DNA 缠绕在"珠"的表面。

电泳的结果和电镜观察到"绳珠"结构之间是什么样的关系呢? Kornberg 和 Thomas 1974 年用实验回答了这一问题。他们先用小球菌核酸酶稍稍消化一下染色质,切断一部分 200 核苷酸对单位之间的 DNA,使其中含有单体、二聚体、三聚体和四聚体等。然后经离心将它们分开。每一组再通过凝胶电泳证明其分子大小及纯度。然后分别用电镜来观察各组的材料;结果单体均为一个 100Å 的小体,二聚体则是两个相联的小体,同样三聚体和四聚体分别由三个小体和四个小体组成,表明 200 核苷酸的电泳片段长度级差正好是电镜观察到的一个"绳珠"单位,他们称其为**核小体**(nucleosome)或核粒,

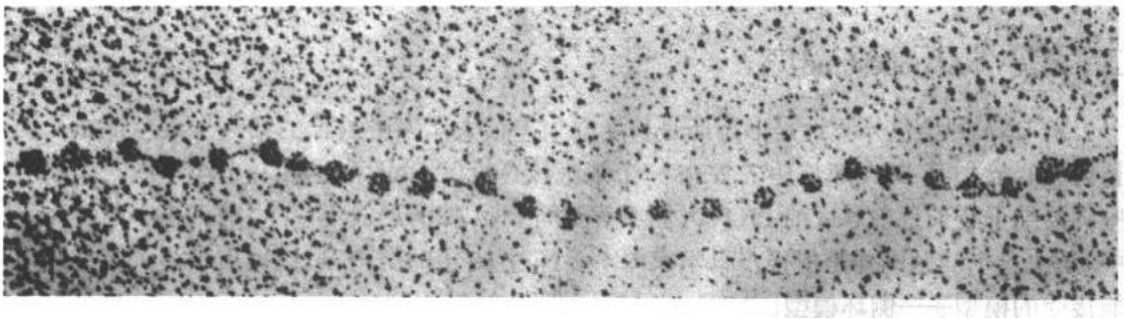


图 10－9　染色质绳－珠结构的电镜照片。(转引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 10－17)

提出了染色质结构的"绳珠"模型。

人们接着用化学交联、高盐分离组蛋白，以及 X 衍射等方法进一步研究组蛋白多聚体的结构、排列以及怎样和 DNA 结合的，从而建立了**核小体模型**。1984 年 Klug 和 Butler 进行了修正。核小体的构造可用图 10－10 表示。

每一个核小体结合的 DNA 总量为 200bp 左右，一般在 150～250 变化范围内。不同生物，不同组织核小体结合 DNA 的长短也不同。此是通过用微球菌核酸酶(micrococcal nuclease)轻微消解染色质而得知的。连接两个核小体的连接 DNA (linker DNA) 是最容易受到这种酶的作用，因此微球菌核酸酶在连接 DNA 处被切断，此时每个重复单位

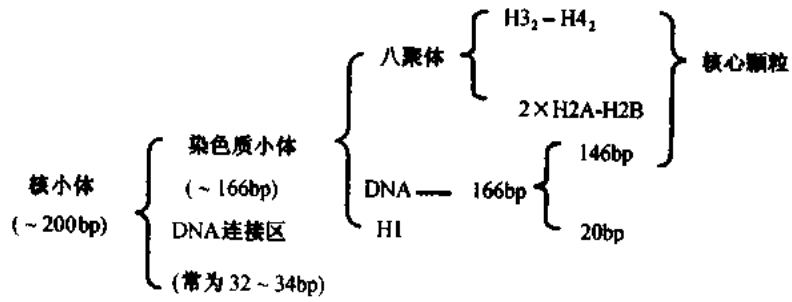


图 10－10　核小体的组成。

的 DNA 长约 200bp，而且是和五种组蛋白相结合，保持着核小体的结构。也就是"绳珠"结构的绳被切断，剩下一个一个的"珠"。

用这种酶进一步消解连接 DNA 的"尾巴"，结果释放出了 H1 留小了核小体核心颗粒(nucleosome core particle)，此是由组蛋白八聚体和四周的长 146bp 的 DNA 组成。此 146bp 的 DNA 因和八聚体组蛋白核心相结合，所以对酶有一定的抗性，故 146bp 这种长度相对比较稳定。而连接 DNA 对酶比较敏忌，其长度变化也比较大，小到 8bp，长到 114bp 不等。实际上在 H1 未释放前，DNA 消解到最短时不是 146bp，而是 166bp，这是因为在 146bp 两端还各有 10bp 的 DNA，从核心颗粒上翘起来，能与 H1 组蛋白结合。H1 像一个搭扣，将绕在八聚体上的 DNA 固定。当将这 20bp 降解掉后，H1 也随之脱落了。在核心颗粒上 146bp 的 DNA 中有 80bp 的 DNA 是紧密地与八聚体结合，正好围绕八聚体的一周(图 10－11)。

除 H1 外四种组蛋白有相互作用形成聚合体的趋势。H3 和 H4 的亲和力最强，首先由两分子的 H3，H4 呈轴对称结合成四聚体；H2A 和 H2B 异二聚体在上下两端和H4

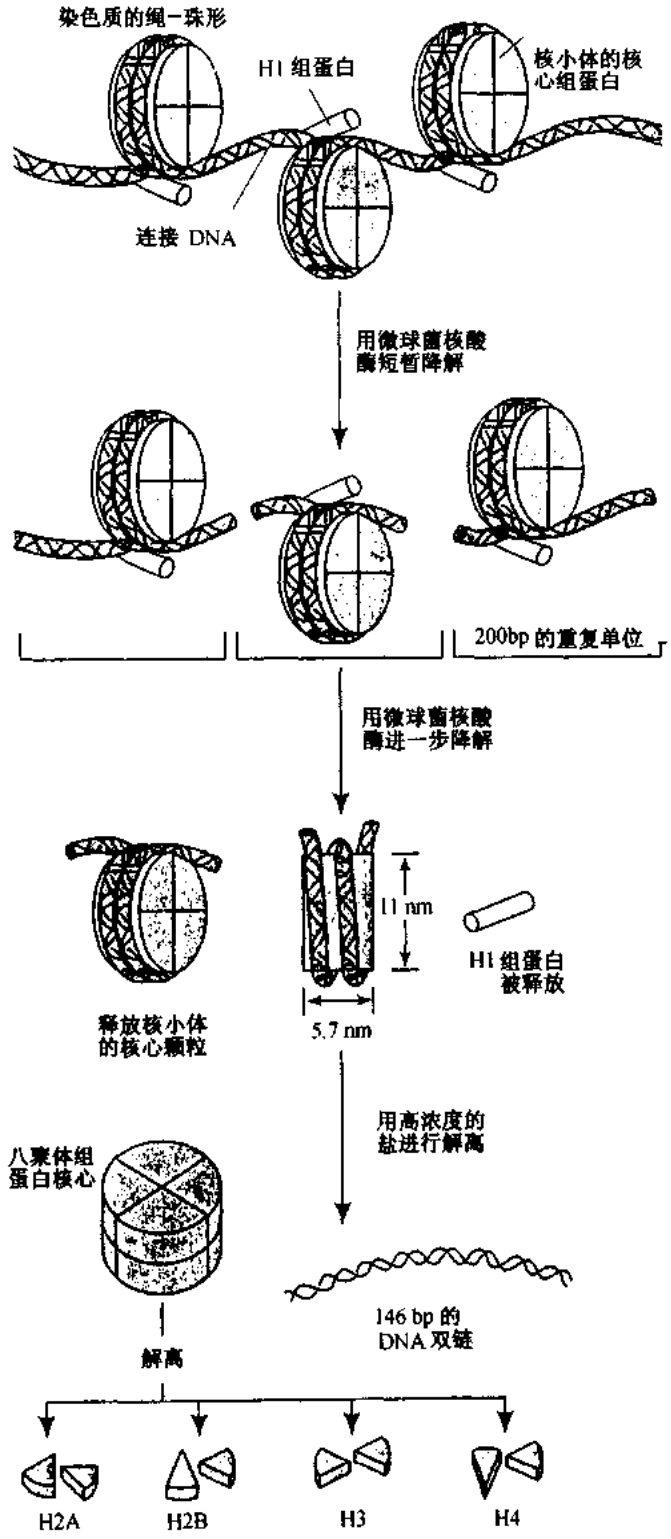


图 10-11 核小体的结构。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 10-18)

的侧面结合形成八聚体(图 10-12)。由于组蛋白八聚体的立体构象形成了沟,沟的位置决定了 DNA 绕在其上的走向为左盘绕,此与 DNA 在溶液中自身环绕成的超螺旋的方向正好相反,然而都是负超螺旋。核小体除去组蛋白之后,DNA 能产生负超螺旋,当

绕在组蛋白上就处于松弛状态。

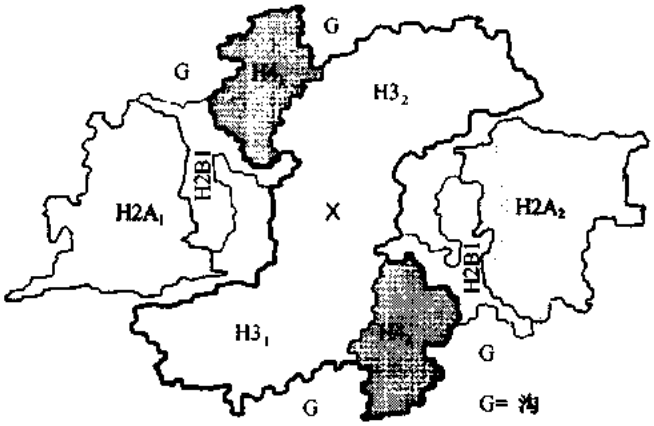


图 10－12　组蛋白八聚体的组成。(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 10－20b)

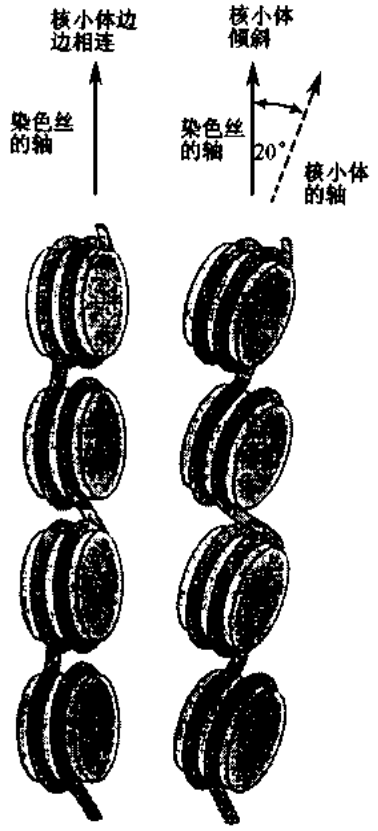


图 10－13　以核小体为基本单位组成了10nm 的染色质丝模型。(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 10－23)

三、染色体的高级结构

在活细胞中染色质并不以“绳珠”状存在。而是保持高度紧密的状态。关于染色质高度折叠的一些细节现在仍不清楚。染色质在细胞周期中发生的一系列规律性变化的细节也不大清楚。相邻核小体彼此紧密相联使染色质形成直径为 10nm 的**核丝** (nucleofilament) (图 10－13)。那么核丝怎样形成更高级的结构呢？主要有两种类型：**螺旋管** (solenoid) **模型和超念珠** (superbeads) 模型。但一般认为螺旋管模型是较为合理的，因为这一模型符合很多作者所观察到的 30nm 的染色体丝(图 10－14)。这一模型是 1976 年 Finch 和 Klug 提出的。他们用小球菌核酸酶消解鼠肝细胞的核，制备了含有 10～100 核小体的染色质，再进行电镜观察。发现在 0.2mmol/L 镁离子存在的条件下 10nm 的核丝进一步螺旋化，形成直径为 30～50nm 的核丝，因而提出 10nm 的染色质纤丝螺旋缠绕成直径为 30nm 的螺旋管，内径为 10nm，螺距 11nm ，螺旋的每一周由 6 个核小体组成(图 10－15)。组蛋白 H1 在形成 30nm 的纤丝的过程必定起到了重要的作用，因为 H1 用完时仍能形成 10 nm 的染色质丝，却不能形成 30nm 的纤丝，但组蛋白 H1 在 30nm 纤丝的什么部位目前尚不清楚。

超念珠模型是由 France(1976 年) 等不同作

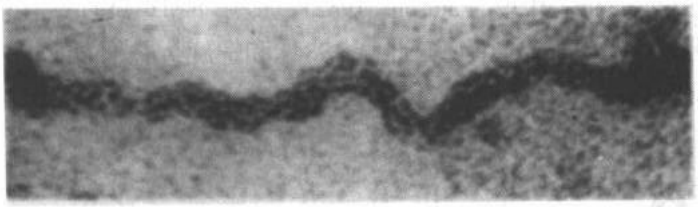


图 10－14　30nm 染色丝的电镜照片。(转引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 10－24)

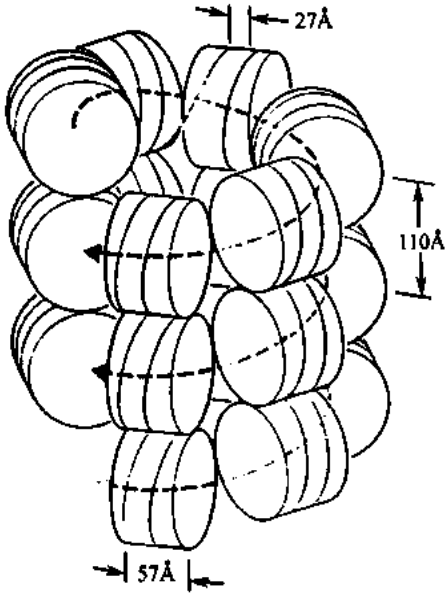


图 10－15　染色质 30nm 螺旋管模型。(转引自 B. Lewin: Genes, Vi 1990, Fig. 21－21)

者提出的。认为染色质的二级结构是由 8～10 个核小体构成的超核小体颗粒。虽然他们在鸡红细胞、海胆精子、玉米分生组织中观察到多个核小体颗粒串联在染色体丝上形成一束一束地排列，但此模型不能解释 30nm 染色质丝的形成。

30nm 的纤丝还不能解决长长的 DNA 如何包装成为染色体，最终纳入小小的细胞核中，也就是说包装比远还没有达到要求。举例来说，人类染色体按 30nm 纤丝的比例的话平均长度有 1mm，而细胞核的直径只有 5μm 左右，染色体的平均长度要比核大 200 倍，怎样能容纳得下 46 条染色体呢？毫无疑问染色质必须进一步地压缩，如何进一步压缩 30nm 的纤丝还不十分清楚。一个间期的染色体直径约为 300nm，而一个中期的染色体直径要达 700nm。

建立中期染色体模型的主要根据是来自以下的资料：Laemmli 等 (1977, 1979) 用 2mol/L 的 NaCl 溶液或硫酸葡聚糖加肝素处理 Hela 细胞中期染色体，除去组蛋白和大部分非组蛋白，在电镜下观察到由非组蛋白构成的染色体骨架 (scaffold) 和由骨架伸出的 30nm 纤丝状 DNA 环状区域 (looped domain)。每一侧环长 15～30μm，相当于 4.5～90kb。(图 10－16)。先用 DNA 酶消解后，再进行脱组蛋白处理仍可得到保持中期染色体形态的非组蛋白骨架。直到再用胰凝乳蛋白酶消化时，此骨架才消失。故骨架确为非组蛋白所组成。由此他们提出了非组蛋白的骨架学说。也有的科学家通过实验认为骨架可能是非组蛋白凝聚产生的假象。但人们对 Laemmli 等揭示的现象仍十分重视。

在骨架模型的基础上人们又先后提出了放射环模型 (radeal loop model) 和螺旋环

模型（Marsden 和 Laemmli，1979，Paulson，1982）、放射环和螺旋环共存模型（Rattner 和 Lin，1985）、袢环模型（J.Painta 和 D.Coffey，1984）和螺旋折叠（helically folded）（Boy de

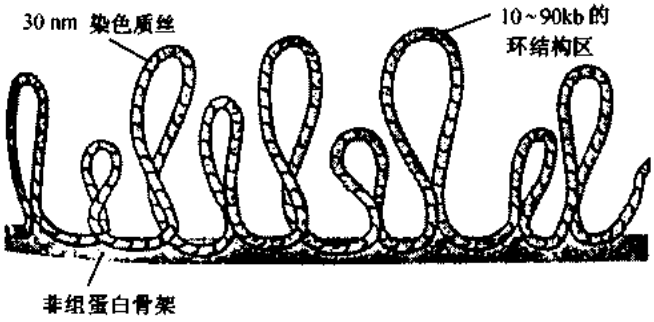


图 10－16　30nm 染色质丝组成的侧环。（转引自 Russell，P.J.：Genetics，3rd ed.，1992，Fig.10－28）

Tour，Ey 和 U.K. Laemmli 1988）等模型，现在一般认为 30nm 的纤丝和非组蛋白骨架结合形成很多侧环，每个侧环长 10～90kb，约 0.5μm，人类染色体约 2000 个环区。（图 10－17）而每个环和染色体的复制单位，以及表达调控有关（见第 19 章）。带有侧环的

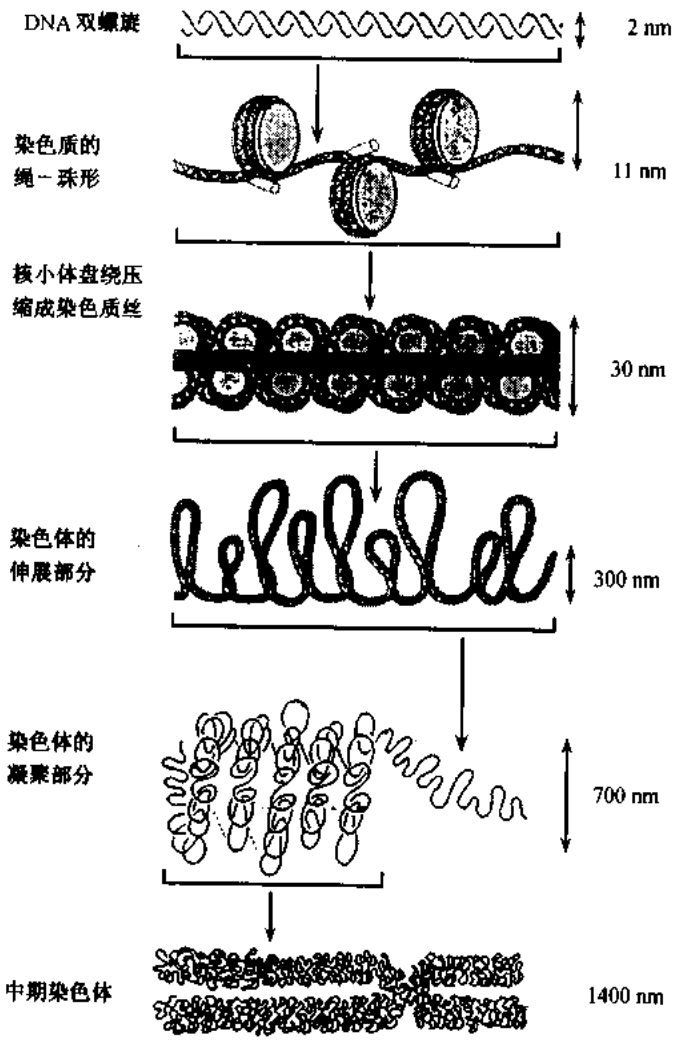


图 10－17　染色体结构的侧环模型。（引自 Russell，P.J.：Genetics，3rd ed.，1992，Fig.10－30）

非组蛋白骨架进一步地形成直径为 700nm 的螺旋，这就是染色单体（图 10－18），而由

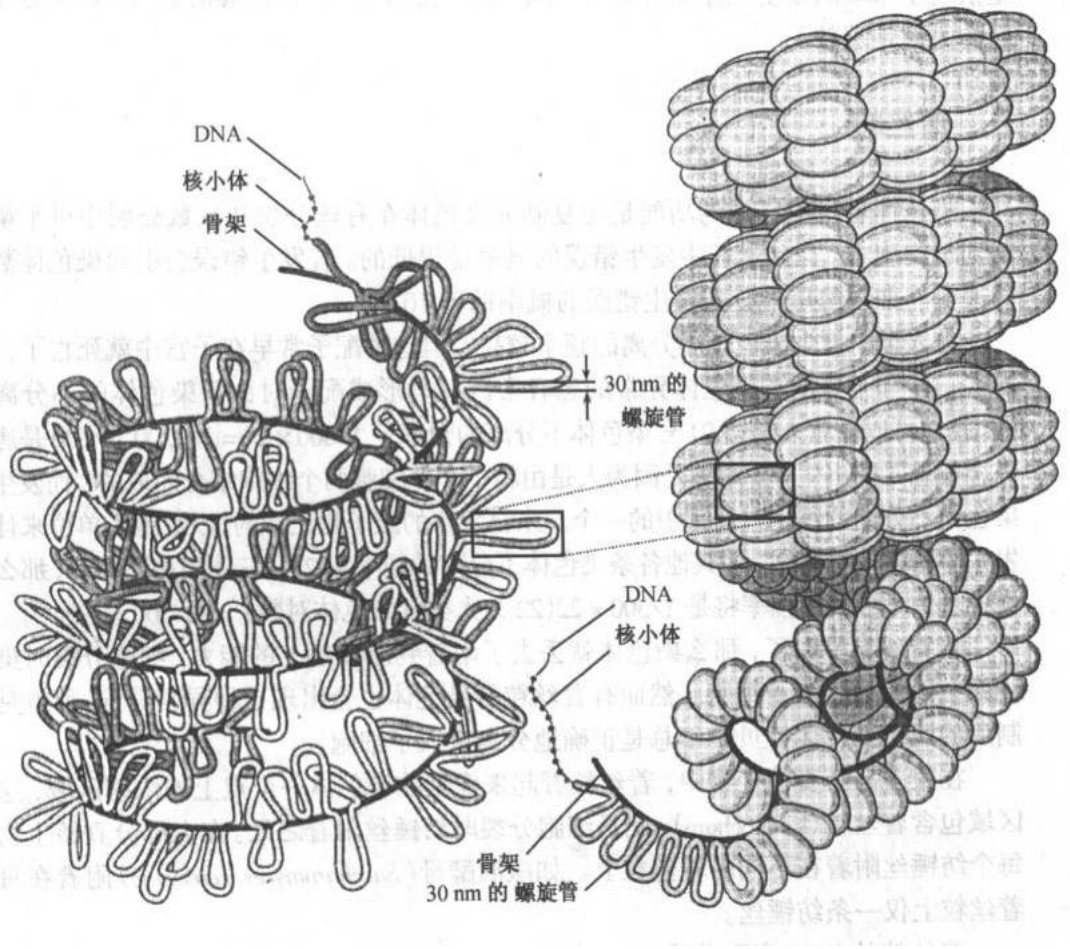


图 10－18　侧环盘绕形成染色单体的模型。（仿 Griffiths A.J.F. *et al*: An Introduction to Genetic Analysis, 5th ed., 1993. Fig. 16－7）

两条姐妹染色单体形成的中期染色体直径为 1400nm，螺旋的方向是相对的。

第三节　着丝粒和端粒

在有丝分裂和减数分裂中复制后的染色体要能均匀准确地分配到两个子细胞中，其关键在于纺锤丝附着在染色体的着丝粒上，使染色体向两极移动。失去了着丝粒的染色体片段就不能精确地分离，也不能和其他的染色体在末期汇聚在一起形成核，而是在细胞质中逐渐凝聚为微核。

在真核基因工程中，我们希望重组在载体上的外源基因不仅能复制，而且随着细胞的分裂，均匀地进入子细胞中，使其具有稳定的遗传性，因此人们在构建真核载体时，设法将着丝粒作为一个构件组装到载体上，如酵母的人工染色体（YAC）就是如此。

每一条染色体的末端都有特殊的结构叫**端粒**（telomere），它可以保持染色体的稳定性，丢失了端粒的染色体很容易和别的染色体相接形成双着丝粒染色体或自身首尾相联形成环状染色体，前者在减数分裂后期会产生**染色体桥**（chromosome bridge），最终引起染色体的重复与缺失（见第 22 章）。端粒还可以使线形真核染色体的复制能顺

利进行，若没有端粒复制后新形成的 DNA 会逐渐减短，在 T7 中，线形 DNA 的复制是依赖于末端的重复顺序来解决这个问题的（见第 11 章），在真核生物中解决这个问题的关键是端粒。

一、着丝粒

染色体着丝粒区域的功能是使复制的染色体在有丝分裂和减数分裂中可平衡分配到子细胞中。在此过程中发生错误的概率是很低的。若发生错误会引起染色体数目的改变。如在酵母中分配发生错误的概率低于 10^{-5}。

在人类中发生染色体不分离的频率较高，但这种配子常早在子宫中就死亡了，像 21－三体和某些 X－染色体异常的患者常因其母形成配子时由于染色体的不分离所致。我们可以估算人类第 21 号染色体不分离的频率是 1/600×2 ＝1/1200。1/600 是唐氏综合征患儿的出生率，乘 2 是因为人是由精子和卵细胞两个细胞结合而成的，而发生21染色体不分离的细胞只是其中的一个。来自父亲的配子是正常的，以细胞为单位来计算发生率所以分母乘 2。若其他各条染色体不分离的频率和 21－三体病相同的话，那么人类染色体不分离的频率将是 1/300×22(22 为人类常染色体对数) ＝22/300＝7%。

若着丝粒丢失了，那么染色体就丢去了附着到纺锤丝上的能力，细胞分裂时染色体就会随机地进入子细胞。然而有着丝粒的染色体也会出现这种异常分配，那就是复制后的两个染色体拷贝并不总是正确地分离进入子细胞。

在很多高等真核生物中，着丝粒看起来像是在染色体一个点上的浓缩区域，这个区域包含着丝粒(kinetochore)。此是细胞分裂时纺锤丝附着之处。在大部分真核生物中每个纺锤丝附着在不同的着丝粒上。如酿酒酵母(*Saccharomyces cerevisiae*)附着在每个着丝粒上仅一条纺锤丝。

着丝粒的 DNA 序列称为 CEN 序列，至少有 10 种酿酒酵母的 CEN 序列已被确定。着丝粒的 DNA 序列有种的特异性，但功能都相同。没有一个已知的 CEN 区和另一个的 DNA 序列及组成完全相同(图 10－19)。但四种酵母的 CEN 区也存在着关键的序

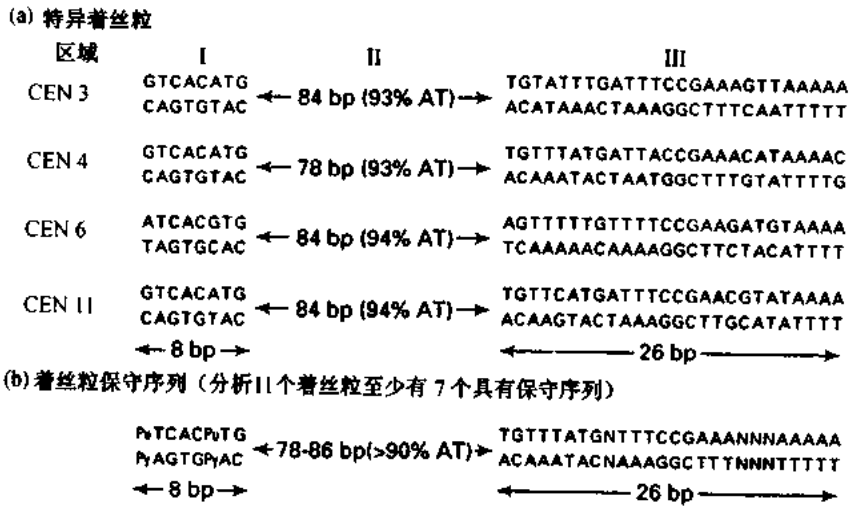


图 10－19　着丝粒 DNA 的特异区域和保守序列。(a)在酿酒酵母的着丝粒 DNA 中 CEN3, CEN4, CEN6 和 CEN11 之间的同源区域。(b)酿酒酵母的保守序列。(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 10－13)

列信息(图 10－19)即保守序列。它们的核心着丝粒区都由三个功能区组成。中部的元件Ⅱ是最大的功能区，由 80～90bp 组成，A＋T 含量高，＞90%；左侧的元件Ⅰ含有 PuTCACPuTG 顺序（Pu 是嘌呤），而右侧的元件Ⅲ是由 26bp 组成的保守区，此区也

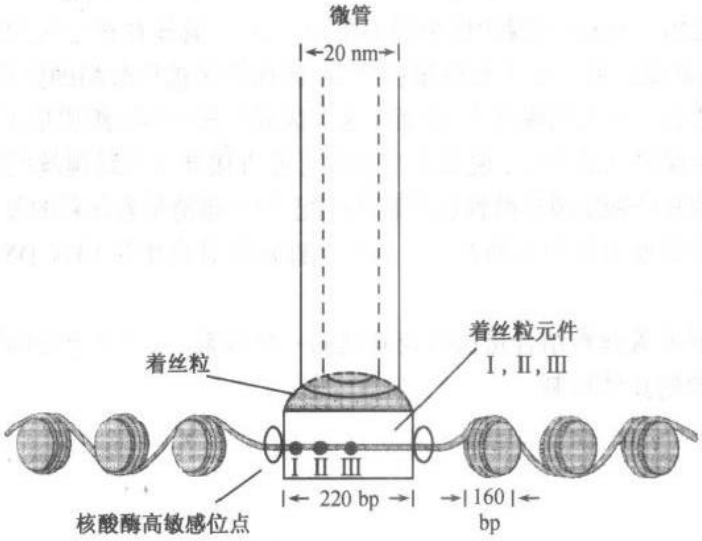


图 10－20　酵母着丝粒结构的模型。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 10－32)

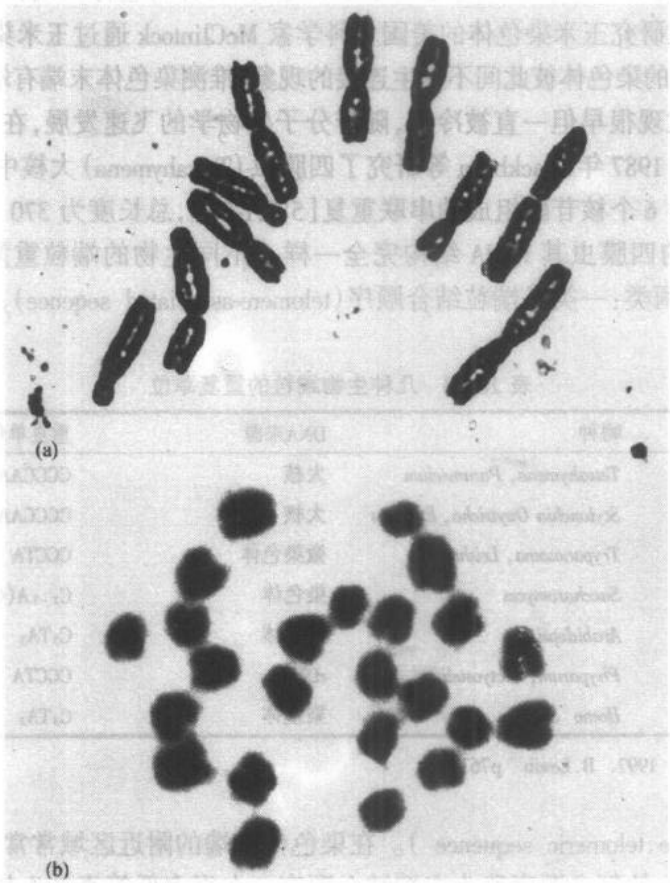


图 10－21　动植物的中期染色体。(a)芍药(*Paeonia lactiflora Pall*)的中期染色体(2n＝10)；(b)黄鳝(*Monopterus albus Zuiew*)的中期染色体(2n＝24)，皆为端着丝粒染色体。(引自徐晋麟等：水生生物学报 1994, 18(2), 164)

是 A－T 丰富区。有趣的是所有真核生物的着丝粒功能都相同，但其 DNA 却具有种的特异性。

啤酒酵母着丝粒的染色质组成已通过染色质位点作图进行了分析。即用对另一种微球菌核酸酶或 DNase Ⅰ的超敏感性来检测其活性。结果表明：着丝粒核心区域Ⅰ、Ⅱ和Ⅲ都被夹在 220～250bp 的保护区中间（图 10－21）。着丝粒核心区的两侧是核酸酶的高敏感区域。着丝粒核心区不和组蛋白结合。此保护区被核酸酶的剪切位点所包围，剪切位点通常在核小体的间隔区上，表面在这个区域中核小体的组成是正常的。CEN 元件Ⅲ是在 220bp 保护区的中心，说明这个元件可能直接涉及到纺锤丝的结合。图中还要注意着丝粒的直径和纺锤丝微管直径非常匹配。另外很清楚着丝粒的复合结构和纺锤体微管的附着必须要有保护区的介入，也许纺锤丝微管直接和 CEN DNA 直接相互作用。

目前正在研究着丝粒结合蛋白以及其他的一些因素。一个主要的问题是解决纺锤丝附着到着丝粒的具体机制。

二、端粒

早在 1938 年著名的遗传学家 Müller 用 X 射线来处理果蝇进行诱变，发现在染色体的末端很少发生缺失和倒位的现象，于是推测在染色体的两端一定存在着一种特殊的结构，这种结构对于染色体的稳定性是至关重要的，并将它称为端粒（telomere）。

同年一直在研究玉米染色体的美国女科学家 McClintock 通过玉米染色体的断裂端容易融合而正常的染色体彼此间不发生连接的现象，推测染色体末端有端粒的存在。

端粒虽然发现很早但一直被冷落，随着分子生物学的飞速发展，在 70 年代的末期终于有了突破。1987 年 Blackburn 等研究了四膜虫（Tetrahymena）大核中的 rDNA 分子，发现其末端是由 6 个核苷酸组成的串联重复 $[5'G_4T_2 3']_n$，总长度为 370～520bp，重复几十次。不同种的四膜虫其 rDNA 结构完全一样。不同生物的端粒重复单位不同（表 10－3），可分为两类：一类是端粒结合顺序（telomere-associated seqence），另一类叫简单

表 10－3　几种生物端粒的重复单位

生物	物种	DNA来源	重复单位(5′－3′)
四膜虫	*Tetrahymena*, *Paramecium*	大核	CCCCAA
尖毛虫	*Stylonchia Oxytricha*, *Euplotes*	大核	CCCCAAAA
鞭毛虫	*Trypanosoma*, *Leishmania*	微染色体	CCCTA
酵母	*Saccharomyces*	染色体	$C_{2-3}A(CA)_{1-3}$
拟南芥菜	*Arabidopsis*	染色体	C_3TA_3
泥毛霉菌	*Phyparun*, *Dictyostelium*	rDNA	CCCTA
人	*Homo Sapiens*	染色体	C_3TA_2

摘自 Genes VI. 1997. B.Lewin p763

端粒顺序（simple telomeric sequence）。在染色体末端的附近区域常常含有重复顺序，但是仍有许多 kb 的复杂顺序称为末端结合顺序。在很多低等真核生物中端粒顺序在染色体 DNA 分子的最末端，或者非常靠近最末端。这些顺序是端粒区重要的功能成分，足以使染色体末端保持稳定。表（11－3）中所列的是在很多低等真核生物中发现的

简单端粒顺序。如四膜虫的端粒顺序从染色体的中部一直延伸到它的末端，这种重复顺序构成的元件是 $\frac{5'CCCCAA-3'}{3'-GGGGTT-5'}$。而在鞭毛虫中重复顺序是 $\frac{5'-TTAGGG-3'}{3'-AATCCC-5'}$。尽管在低等真核生物中，简单的端粒顺序多种多样，但都有以下的共同特点：

(1) 每一个端粒都含有一系列的短的正向重复顺序(表 10-3)。这些顺序都可用 $C_n(A/T)_m$ 这一通式来表示，其中 $n>1$，而 $m=1\sim4$。

(2)在端粒区域中有一种特殊的不连续排列，产生单链断裂的形式，这种结构并不能被连接酶将缺口封闭起来，而在正常情况下连接酶是可以作用 DNA 链上的缺刻(nick)。这表明这种断裂所形成的缺口较大。因为连接酶只能连接相邻的两个 3′-OH 和 5′-p 末端之间形成磷酸二酯键，而不能封闭大的裂缺(gap)。提纯的天然的端粒区可直接用 *E.coli* DNA 聚合酶Ⅰ进行缺口平移(nick translation)(见 13 章)，表明在端粒区确有缺口存在，缺口中的 3′-OH 为缺口平移时新链的合成提供了引物。

(3)在端粒区的最末端以某种方式被封闭起来，可能带有 3′-OH 的单链末端回折成了发夹回(hairpin)结构，因此它们不被核酸酶所识别。1987 年 E.Henderson 等用非变性聚丙烯酰胺凝胶电泳，吸热变性分析以及核磁共振光谱分析了端粒的顶端结构，证实了端粒顶端回折形成发夹结构，而这一结构并不是由通常的 Watson-Crick 配对，而是通过 G-G 配对而形成，富含 G 的特点在形成特殊的 G-G 配对及染色体的复制有着重要意义。

(4) 端粒的双链部分中含有 T_2G_4 的顺序在 3′末端，排列方向是朝向染色体的中心部位。

(5)在细胞中，端粒总是和一些非组蛋白相结合，形成复合体，常表现为异染色质或构成染色体的末端结节，所以我们在对染色体处理后进行硝酸银染色时，发现端粒部位出现浓染，其他部位出现浅染或不染色。D.E.Gottschling 和 K.A.Zakian(1986 年)在尖毛虫(dxytricha)中发现两种端粒蛋白，分子量分别为 55kDa，26kDa。它们与端粒 DNA 紧密结合，能抗高盐提取和核酸酶的处理。他们推测这种 DNA 与蛋白的结合是非共价结合，可能是依赖氢键和范特华尔力的作用。人们进一步发现在体外这种端粒蛋白与端粒的重复顺序也能形成复合体。

端粒除了保持染色体的稳定和线形 DNA 的顺利复制以外，随着研究的逐步深入，人们又发现它还有一些其他的生物功能。如和染色体的行为、细胞的寿命、遗传信息的复制和表达以及核骨架的组成都有一定的关系。

人们发现无论是同源染色体还是异源染色体，它们的端粒常发生暂时性的并合。在第一次减数分裂前期，端粒聚集在一起并处于核膜的周围，使染色质呈“花束”(bouquet orientation)状排列。这种端粒与端粒之间的互相作用的意义还不太清楚，但人们认为可能影响到染色体的行为如重组等。

近年来(20 世纪 90 年代前期)的研究证明生殖细胞可以重建与延长端粒的 DNA；而体细胞则随着复制次数的增加端粒 DNA 会逐渐缩短。Harley 等(1989 年)用人的端粒重复片段为探针对胎儿、新生儿、青年和老年人的细胞株的端粒长度进行了研究，发现随着年龄的增加，人成纤维细胞端粒的长度不断下降。因此端粒 DNA 可能决定了细胞的寿命，并与细胞的自然凋亡与癌变密切相关。1990 年 Harely 等则认为端粒的长度可反映细胞有丝分裂的次数。最有力的证据是分离培养 0 岁(胚胎组织)~93 岁不同年龄人的成纤维细胞，其端粒的长度与细胞分裂有关，细胞分裂的次数与细胞最初的端粒长度成正比。进一步研究发现早老性侏儒症患者细胞端粒明

显较正常人为短，其细胞在培养中分裂能力减低，以上表明端粒像是细胞分裂的“计数器”。

早在 20 世纪 60 年代和 70 年代人们就观察到端粒和着丝粒均靠近核膜并推测这种与核膜的紧密结合关系，维持染色体的空间定向，使染色体所携带的信息能有序而有效地复制和表达。

翟中和等为了研究端粒与核骨架的关系用 SDS－聚丙烯酰胺凝胶双向电泳分析核骨架的蛋白组成，然后经 Western 转移，再与同位素标记的端粒 DNA 探针一起杂交，结果表明端粒 DNA 序列与核纤层蛋白(lamin)、波形蛋白(vimentin)以及某些核骨架蛋白质有亲和关系。Qiu 和 Henderson(1991 年)提纯并鉴定了四膜虫的端粒蛋白，他们发现端粒蛋白和核骨架结合。因此翟中和等认为端粒蛋白本身可能就是核骨架的组成之一。

第四节　染色体的核型和分带

原核生物和真核生物之间根本的区别是原核生物只有一条染色体，而大部分真核生物几乎所有的体细胞中都具有二倍体数目的染色体，即一般为两套染色体。在间期核中用光镜是看不到染色体的。在有丝分裂中染色体经染色后其形态和数目在光镜下就可以观察得十分清楚(图 10－21)。中期染色体每一个都具有两个姐妹染色单体，它们都处于高度凝聚的状态。一对姐妹染色单体依赖着丝点彼此联在一起。几乎所有的细胞遗传学的分析都要依赖中期染色体。图 10－22 表示人类男性的中期染色体，共有 46 条，每一条都呈 X 形，它们的大小和着丝点的位置都不同。

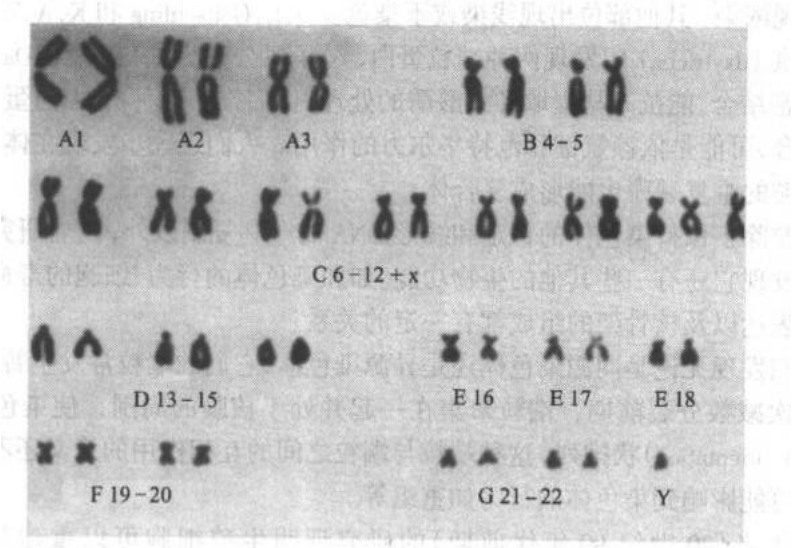


图 10－22　正常男人的核型。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 10－13)

一、核型

在一个细胞中全套的中期染色体称为核型(Karyotype),或染色体组型。对于大部分生物来说个体所有细胞的核型都相同,但却有种的特异性。在真核生物中染色体的数目、大小和形态有一个很大的范围。甚至亲缘关系很近的生物它们的核型可能完全不同。表 11-4 为我们提供了一些普通生物的染色体数目。人类二倍体染色体数目是46,而黑猩猩是 48,猫是 38,豌豆是 14,黑腹果蝇是 8。在单倍体生物如衣藻中是 16,在酿酒酵母中是 17。图 10-22 是正常男人的核型。这是将人类中期染色体照片按大小排列而成。此核型表示 46 条染色体,其中 22 对常染色体可以配对,而 X 和 Y 染色体大小和形态都不同。

染色体的大小、形态和带型(banding patterns)的资料可以用来分析与先天异常和功能紊乱有关的染色体畸变,例如较为罕见的染色体增加了一个片段或丢失了一个片

表 10-4 一些生物的染色体数目

生 物	2N	生 物	2N
人类(*Homo sapiens*)	46	水螅(*Hydra vulgaris attenuata*)	32
黑猩猩(*Pan troglodytes*)	48	海绵(*Sycon eiliatum*)	26
金丝猴(*Rhinopithecus roxellanae*)	44	水绵(*Spirogyrasp*)	24
猕猴(*Macaca malatta*)	42	海带(*Laminaria japonica*)	44
马(*Equus caballus*)	64	蕨(*Pteridium aquilinum*)	104
驴(*Equus asinus*)	62	银杏(*Ginkgo biloba*)	24
黄牛(*Bos taurus*)	60	黑松(*Pinus thunbergii*)	24
水牛(*Bobalus buffelus*)	48	银白杨(*Populus alba*)	38
绵羊(*Ovis aries*)	54	桑(*Morus alba*)	28
山羊(*Capara hircus*)	60	沙梨(*Pyrus alba*)	34
狗(*Canis familiaris*)	78	陆地棉(*Gossypium hirsutum*)	52
猪(*Sus scrofa*)	38	中棉(*G. arboreum*)	26
猫(*Felis catus*)	38	大麦(*Hordeum vulgare*)	14
兔(*Oryctolagus cuniculus*)	44	水稻(*Oryza sativa*)	24
豚鼠(*Cavia cobaya*)	64	玉米(*Zea mays*)	20
大鼠(*Rattus norvegieus*)	42	一粒小麦(*Triticum monococcum*)	14
小鼠(*Mus musculus*)	40	普通小麦(*T. aestivum*)	42
火鸡(*meleagris gallopavo*)	82	洋葱(*Allium monococcum*)	16
家鸡(*Gallus domesticus*)	78	黄瓜(*Cucumis sativus*)	14
鸭(*Anao platyrhytcha*)	80	西瓜(*citrullus vulgaris*)	22
青蛙(*Rana nigromaculata*)	26	碗豆(*Pisum sativum*)	14
蟾蜍(*Bufo bufo*)	22	香碗豆(*Lathyrus adoratus*)	14
爪蟾(*Xenopus laevis*)	36	蚕豆(*Vicia faba*)	12
蝾螈(*Salamandra salanandra*)	24	菜豆(*Phaseolus vulgaris*)	22
鲤鱼(*Gyprinus carpio*)	100,104	金鱼草(*Antirhinum mayus*)	16
文昌鱼(*Amphioxus lanceolatus*)	24	向日葵(*Antirrhinum mayus*)	34
蚯蚓(*Perichaeta communissima*)	32	大白菜(*Brassica pekinensis*)	20
线虫(*Caenorhabditis elegants*)	11雄 12雌	甘蓝(*B - oleracea*)	18
马蛔虫(*Ascaris megalocephalus*)	4	烟草(*Nicotiana tabacum*)	48
家蚕(*Musca domestica*)	56	番茄(*Salanum lycopersicum*)	24
家蝇(*Musca domestica*)	12	藻类和真菌	N
果蝇(*Drosophila melanogaster*)	8	衣藻(*Chlamydomonas reinhatt*)	16
蜜蜂(*Apis mellifera*)	32雌 16雄	酵母(*Saccharomyces cerevisiae*)	17
蚊(*Culex pipiens*)	6	曲霉(*Aspergillus nadulans*)	8
佛蝗(*Phlaeoba infumata*)	24雌 23雄	链孢霉(*Neurospora crassa*)	7
		青霉菌(*Penicillium species*)	4

段或者非同源染色体之间交换了一个片段；或出现一些异常的环形染色体、双着丝点染色体、等臂染色体（isochromosome）、异常的端着丝点染色体或额外的小染色体（extra small chroimosome）。另外就是染色体数目的改变，由于在配子形成过程中，细胞分裂时染色体的不分离现象会导致染色体数目的增加和减少。如额外多了一条第21号染色体而导致产生先天性愚型或称唐氏综合征（Down's syndrome）。第5号染色体短臂的缺失也会产生猫叫综合征（cri-duchat syndrome 或 cat cry syndrome）。

在人类的核型中，染色体数目是很容易确定的。最大的染色体只有一对（A_1），次大的也只有一对（A_2），其余都是中等的和小的染色体，从1号到22号都是常染色体。X染色体是一个中等大小的中央着丝粒染色体。Y染色体是很小的染色体。

图10-23　雌性小鼠的染色体G带核型。（转引自 Griffiths A. J. F. *et al*.: An Introduction to Genetic Analysis. 5th ed., 1993. Fig. 6-25b）

正常的动、植物核型还可以作为研究分类学和系统学的重要依据。经典分类学主要是依据动植物的各种形态特征。在植物中常依赖的是根、茎、叶、花和果实的形态和结构，花的形态结构尤为重要，由于生物形态特征在不同的生态环境中会产生一定的变异，因此在分类中常引起一些困惑和争论，从而在80年代以来又形成了一个新的分支叫细胞分类学，也就是利用各种植物的正常核型的特异性和系统进化中的相关性，来研究形态分类和系统学中的一些疑难问题。比如百合科植物黄精（Polygonatum）约有40余种。但黄精的分类正如《中国植物》中所指出的那样："黄精某些种之间在一定形态上既有区别，又有交错；在地理分布上，也是既有替代现象，又有重叠之处。因此，本属种的划分是一个十分困难的问题"。Baker（1875）和 Satake（1942）都试图从叶序和苞片大小特征来进一步划分，但都没有得到满意的结果。从20世纪30年代起直至现在国

内外很多学者都用核型分析来研究黄精属内种的划分，提供了很多有用的资料。

二、染色体带型

一般的染色体核型常常会遇到三种难题：①是由于在制片过程中染色体的大小和形态会发生一些改变，特别是在染色体大小和形态相近的不同号数的染色体之间难以鉴别，如人类的C组（从第6对到第12对）染色体就难以准确鉴别，有时寻找一对同源染色体也较困难，只好凭着主观的判断，因此"错点鸳鸯谱"的现象也很难避免；②难以确定重复、缺失和倒位的存在，更难鉴别其类型，如是相互易位还是单向易位，是末端易位还是中间易位等等。三是染色体没有一定的界标，在基因定位时难以描述，比如红绿色盲的基因定位在染色体的长臂上，但具体的位置如何描述才能使人一看就知道其座位及和其他基因的关系呢？为了解决这些问题，人们又发展了一项技术那就是染色体的分带染色。分带染色技术是将未染色的中期染色体片经过一定的预处理，再用不同的方法染色，使染色体上出现明显而稳定的染色条带。每一号染色体都有不同的带型，这样以上三大难题也就迎刃而解了。分带技术由于染色与预处理的方式不同可产生不同的带型，因此有G带、C带、Q带、N带和R带等不同技术。

G带技术是其中最常用的技术，由Pardue和Gall（1970）建立。方法是先用加热或蛋白水解酶处理未染色的中期染色体片，然后用姬姆萨（Giemsa）染料染色，结果染色体呈现清晰、特异的G带（G banding）（图10-23）。G带反映了染色体DNA上A-T的丰富区，在人类中约有2000条G带可被鉴别。G带不仅稳定且用扫描电镜可在带区见到收缩的痕迹（图10-24）。现在已制成人类的G带标准图谱（图10-25）。可用于鉴定染

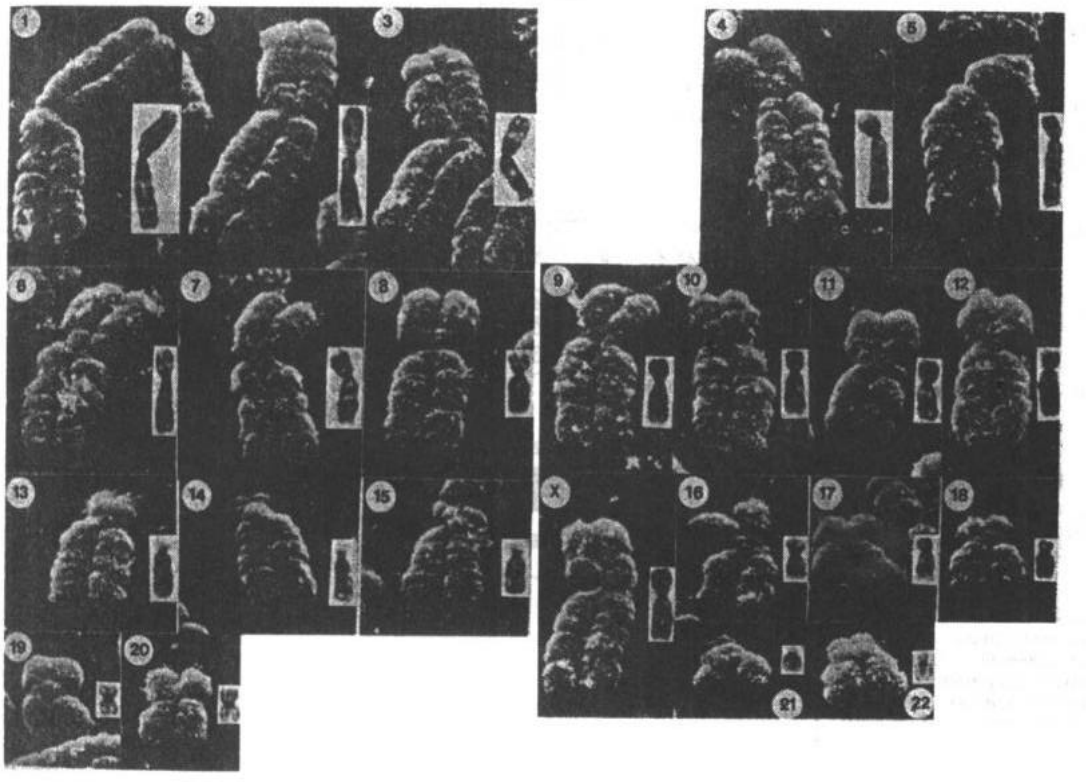


图10-24　人类染色体C带的形态。在扫描电镜下可见到带区有收缩的痕迹。（引自Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 10-15）

色体号数以及基因定位。在植物中 G 分带技术尚不成熟，不少植物中难以显示 G 带或者结果很不稳定。

C 带技术是和 G 带技术同时建立的，主要用以显示染色体中的组成型异染区，如着丝粒带(图 10－26)。常用于植物的染色体分类。

Q 带是在 1968 年由 Caspersson 等建立的。这种方法是将中期染色体用芥子喹吖因(guinacrine mustard)染色，在染色体的 G 带相同区域产生另外荧光带，在荧光显微镜下可以清晰地看到(图 10－27)。Q 带虽然可以用于诊断，但受到荧光染色所限，不能长

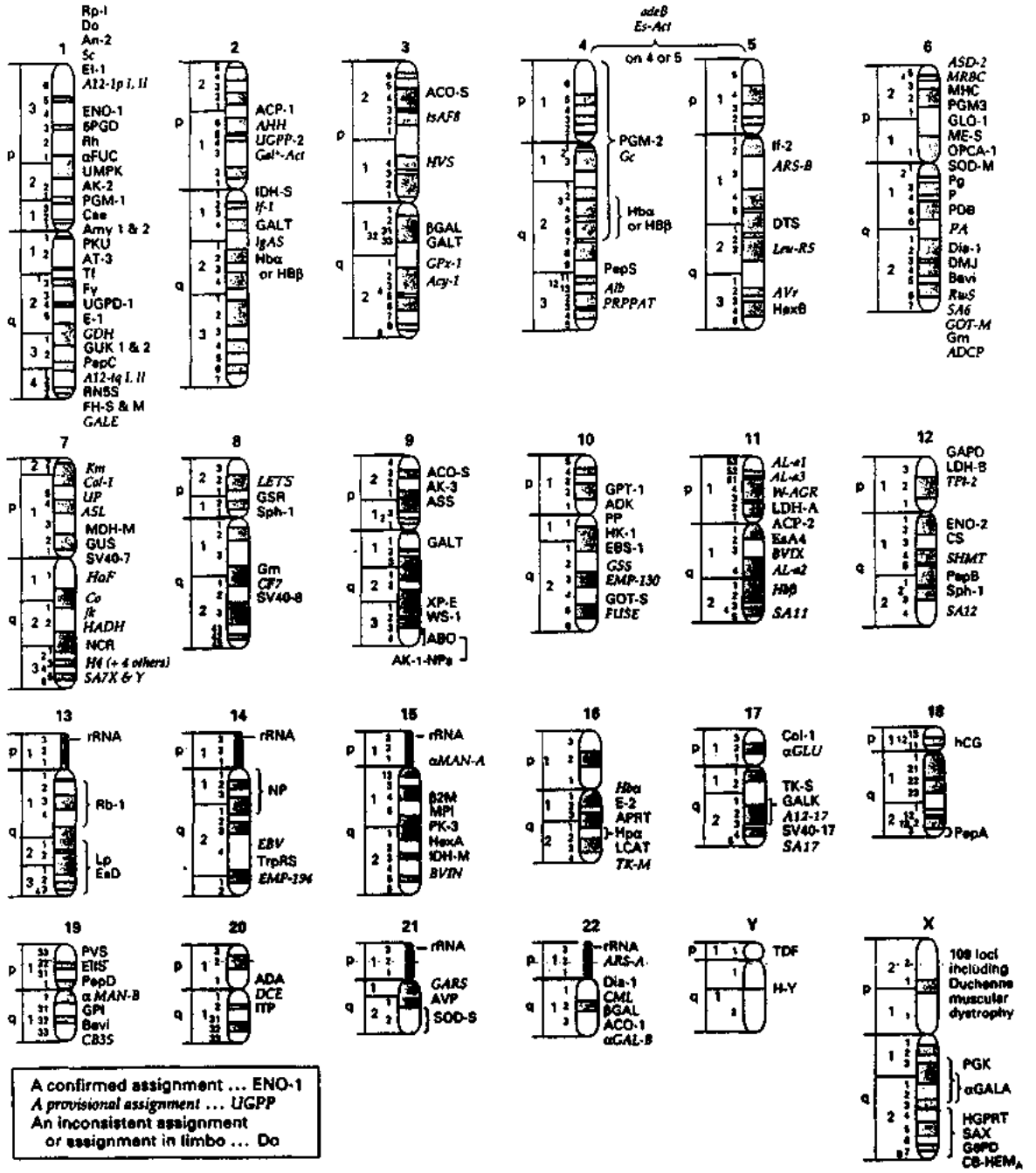


图 10－25　人类 G 带的标准图谱。(引自 Russell, P. J.: Genetics, 3 rd ed., 1992, Fig. 6－20)(图中的关键基因见下页表)

图 10-25 中的相关基因

基因符号	产物和性状	定位(No)
ABO(ABO blod group)	ABO 血型	9
ACO-M(Acomitase, mitochondrial)	线粒体乌头酸酶	22
ACO-S(Acomitase,soluble)	可溶性乌头酸酶	9
ACP-1(Acid phosphatase -1)	酸性磷酸酶 1	2
ACP-2(Acid phosphatase -2)	酸性磷酸酶 2	11
Acy-1(aminoacylase-1)	酰化氨基酸水解酶	3
AdeB(Formylglycinamide ribotide amidotransferase)	甲酰甘氨酸酰核苷酸酰胺转移酶	4 或 5
ADA(Adenosine deaminase)	腺苷脱氨酶	20
ADCP(Adenosine deaminase complexingprotein)	腺苷脱氨酶复合蛋白	6
ADK(Adenosine kinase)	腺苷激酶	10
A12-1pI,II(Adenovirus-12 chromosome modification site-1pIandII)	腺病毒-12 染色体修饰位点 1pI和II	1p
A12-1qI,II(Adenovirus-12 chromosome modification site-1qIandII)	腺病毒-12 染色体修饰位点 1pⅠ和Ⅱ	1q
A12-17(Adenovirus-12 chromosome modification site-17)	腺病毒-12 染色体修饰位点 17	17
AH-3(Adrenal hyperplasia or 21-hydroxylase deficiency)	肾上腺增生或 21 羟化酶缺陷	6
AHH(Aryl hydrocarbon hydroxylase)	芳香烃羟化酶	2
Ak-1(Adenylate kinase-1)	腺苷酸激酶 1	9
Ak-2(Adenylate kinase-2)	腺苷酸激酶 2	1
Ak-3(Adenylate kinase-3)	腺苷酸激酶 3	9
AL(Lethal antigen 3 loci:a1、a2、a3)	致死抗原 3 个位点:a1、a2、a3	11
Alb(Albumin)	白蛋白	4
Amy-1(Amylase,salivary)	唾液淀粉酶-1	1
Amy-2(腺苷酸激酶 1,pancreatic)	胰腺淀粉酶	1
An-2(Aniridia,type Ⅱ Baltimore)	巴的摩尔Ⅱ型无虹膜症	1
ARS-A(Arylsulfatase A)	芳基硫酸酯酶 A	22
ARS-B(Arylsulfatase B)	芳基硫酸酯酶 B	5
APRT(Adenine phosphoriosyltransferase)	腺嘌呤磷酸核糖转移酶	16
ASD-2(Atrial septal defect,secundum type)	继发型房间隔缺损	6
ASL(Argininosuccinate lyase)	精氨琥珀酸裂解酶	7
ASS(Argininosuccinate synthetase)	精氨琥珀酸合成酶	9
AT-3(Antithrombin Ⅲ)	抗凝血酶Ⅲ	1
AVP(Antiviral proten)	抗病毒蛋白	21
Avr(Antiviral state regulation)	抗病毒状态调节剂	5
Bevi(Baboon M7 virus infection)	狒狒 M7 病毒感染	6 或 19
Bf(Properdin factor B)	备解素因子	6
β2 M(β2-microglobulin)	β2 微球蛋白	15
BVIN(BALB virus induction,N-tropic)	嗜神经性病毒诱导	15
BVIX(BALB virus induction xenotropic)	嗜异性病毒诱导	11
C2(Complement component-2)	补体成分 2	6
C4F(Complement component-4 fast)	补体快成分 2	6
C4S(Complement component-4 slow)	补体慢成分 4	6
C6(Complement component-6)	补体成分 6	6
C8(Complement component-8)	补体成分 8	6
Cae(Cataract, zonular pulverulent)	小带状粉末样白内障	1
CB(Color blindness(deutan and protein))	红绿色盲	X
CB3S(Coxsackie B3 virus susceptibility)	柯萨奇 B3 病毒易感性	19
CF7(Clotting factor Ⅶ)	凝血因子Ⅶ	8
Ch(Chido blood group-same as C4S)	Chido 血型	6
CML(Chronic myeloid leukemia)	慢性粒细胞白血病	22
Co(Colton blood group)	Colton 血型	7
Col-1(Collagen 1 (α-1 and α-2)	胶原Ⅰ(α-1 和 α-2)	7 和 17
CS(Citrate synthase, mitochondrial)	线粒体的柠檬酸合成酶	12
Dia-1(NADH-diaphorase)	还原型辅酶-硫辛酰胺脱氢酶	6 或 22

续表

DMJ(Juvenile diabetes mellitus)	幼儿糖尿病	6
Do(Dombrock blood group)	Dombrock 血型	1
DCE(Desmosterol - to - cholisterol enzyme)	24 脱氢胆固醇 - 胆固醇酶	20
DTS(Diphtheria toxin sensivity)	白喉毒素敏感性	5
EBS - 1(Epidermolysisbullosa, Ogna type)	Ogna 型表皮水泡症	10
EBV(Epstein - Barr virus integration site)	EB 病毒整合位点	14
E - 1(Pseudocholinesterase - 1)	拟胆碱酯酶 - 1	1
E - 2(Pseudocholinesterase - 2)	拟胆碱酯酶 - 2	16
E11 - S(Echo 11 sensitivity)	Echo 11 敏感性	19
El - 1(Elliptocytosis - 1)	椭圆形红细胞增多症	1
EMP - 130(External membrane protein - 130)	外膜蛋白 - 130	10
EMP - 195(External membrane protein - 195)	外膜蛋白 - 195	14
ENO - 1(Enolase - 1)	烯醇化酶 - 1	1
ENO - 2(Enolase - 2)	烯醇化酶 - 2	12
Es - Act(Esterase activator)	酯酶激活因子	4 或 5
EsA4(Esterase - A4)	酯酶 - A4	11
EsD(Esterase D)	酯酶 D	13
FH - M(Fumarate hydatase, mitochondrial)	线粒体的延胡索酸水合酶	1
FH - S(Fumarate hydatase, soluble)	可溶性的延胡索酸水合酶	1
αFUC(Alipha - L - fucoidase)	α - L - 岩藻糖 苷酶	1
FUSE(Polykaryocytosis inducer)	多核细胞增多症诱导物	10
Fy(Duffy blood group)	Duffy 血型	1
Gal + - Act(Galactose + Activator)	半乳糖 + 激活因子	2
αGal A(α - galactosidase A(Fabry disease))	α - 半乳糖苷酶 A(费勃莱氏症)	X
αGal B(α - galactosidase B)	α - 半乳糖苷酶 B	22
βGal B(β - galactosidase)	β - 半乳糖苷酶	3 或 12,22
GALK(Galactokinase)	半乳糖苷激酶	17
GALT(Galactose - 1 - phosphate uridyltransferase)	半乳糖 - 1 - 磷酸尿苷酰转移酶	2,3 或 9
GALE(Galactokinase - 4 - epimerase)	半乳糖 - 4 - 表面异构酶	1
αGLU(α - glucosidase)	α - 葡萄糖苷酶	17
GAPD(Glyceraldehyde - 3 - phosphate dehydrogenase)	甘油醛 - 3 - 磷酸脱氢酶	12
GAPS(Phosphoribosyl glucinamide synthetase)	磷酸核糖甘氨酰胺合成酶	21
Gc(Group - specific component)	血型特性成分	4
GDH(Glucose dehydrogenase)	葡萄糖脱氢酶	1
GLO - 1(Glyoxylase 1)	乙二醛酶 1	6
Gm(Immunoglobulin heavy)	免疫球蛋白重链	8
GOT - M(Glutamate oxaloacetate transminase, mitochondrial)	线粒体的谷氨酰草酰乙酸转氨酶	6
GOT - S(Glutamate oxaloacetate transminase, soluble)	可溶性的谷氨酰草酰乙酸转氨酶	10
GPI(Glucosephosphate isomerase)	葡萄糖磷酸异构酶	19
GPT - 1(Glutamate pyruvate transaminase, soluble)	可溶性的谷氨酸丙酮酸转氨酶	10
GPx - 1(Glutathione peroxidase - 1)	谷胱甘肽过氧化物酶 1	3
G6PD(Glucose - 6 - phosphate dehydrogenase)	葡萄糖 - 6 - 脱氢酶	X
GSR(Gluthione reductase)	谷胱甘肽还原酶	8
GSS(Glutamate - γ - semialdehyde synthetase)	谷氨酸 - γ - 半醛合成酶	10
GARS(Glycinamide ribonucleotide synthtase)	甘氨酰胺核苷酸合成酶	21
GUK - 1&2(Guanylate kinase - 1 & 2)	鸟苷酸激酶 1 和 2	1
GUS(β - glucuronidase)	β - 葡萄糖苷酸酶	7
H4(Histone H4 and 4 other histone genes)	组蛋白 H4 和其它组蛋白基因	7
HADH(Hydroxyaacyl - CoA dehydrogenase)	羟酰辅酶 A 脱氢酶	7
HaF(Hageman factor)	凝血因子Ⅻ, Hageman 因子	7
Hbα(Hemoglobin alpha chain)	血红蛋白 α 链	2,4,5,6 或 7
Hch(hemochromatosis)	色素性肝硬变	6

续表

HEM－A(Classic hemophilia)	血友病 A,第Ⅷ因子	X
HexA(Hexosaminidase A)	氨基己糖酶 A	15
HexB(Hexosaminidase B)	氨基己糖酶 B	5
HGPRT(Hypoxanthine－guanine phosphoribosyltransferase)	次黄嘌呤－鸟嘌呤磷酸核糖转移酶	X
HK－1(Hexokinase－1)	己糖激酶	10
HLA(A－D)(Human leukocyte antigens)	人类白细胞抗原	6
HLA－DR(Human leukocyte antigens,D－related)	D－相关人类白细胞抗原	6
Hpα(Haptoglobin,alpha)	α结合珠蛋白	16
HVS(Herpes virus sensitivity)	疱疹病毒敏感性	6
H－Y(Y histocompatility antigen)	Y 组织相容性抗原	Y
IgAS(Immunoglobulin heavy chains attachment site)	免疫球蛋白重链附着位点	2
If－1(Interferon－1)	干扰素－1	2
If－2(Interferon－2)	干扰素－2	5
IDH－M(Isocitrate dehydrogenase, mitochondrial)	线粒体异柠檬酸脱氢酶	15
IDH－S(Isocitrate dehydrogenase,soluble)	可溶性异柠檬酸脱氢酶	2
ITP(Inosine triphosphatase)	次黄苷三磷酸化酶	20
Jk(Kidd blood group)	Kidd 血型	7
Km(Kappa immunoglobulin light chains,Inv)	免疫球蛋白轻链基因簇	7
LAP(Laryngeal adductor paralysis)	喉内收肌麻痹	6
LCAT(Lecithin－cholesterol acyltransferase)	卵磷脂－胆固醇酰基转移酶	16
LDH－A(Lactate dehydrogenase A)	乳酸脱氢酶 A	11
LDH－B(Lactate dehydrogenase B)	乳酸脱氢酶 B	12
LETS(Large, external, transformation－sensitive protein)	外部大的转化敏感蛋白	8
LeuRS(Leucyl－tRNA synthetase)	亮氨酰－tRNA 合成酶	5
Lp(Lipoprotein β－Lp)	脂蛋白	13
AMAN-A(Cytoplasmic α－D－mannosidase)	胞浆的－α－D 甘露糖苷酶 A	15
AMAN-B(Cytoplasmic α－D－mannosidase)	胞浆的－α－D 甘露糖苷酶 B	19
MDH－M(Malate dehydrogenase,mitochondrial)	线粒体的苹果酸脱氢酶	7
MDH－M(Malate dehydrogenase,soluble)	可溶性的苹果酸脱氢酶	2
ME－S(Malic enzyme,soluble)	可溶性的苹果酸酶	6
MHC(Major histocompatibility complex)	主要组织相容性复合物	6
MLC－W(Mixed lymphocyte culture,weak)	弱的混合淋巴细胞培养	6
MPI(Mannosephosphate isomerase)	甘露糖磷酸异构酶	15
MRBC(Monkey red blood cell receptor)	猴红细胞受体	6
MTR(5－Methyltetrahydrofolate:L－homocysteine S－S methyl-transferase)	5-Me-TH4,L-同型半胱氨酸 S-S Me-转移酶	1
NCR(Neutrophil chemotactic response)	嗜中性细胞趋化反应	7
NDF(Neutrophil differentiation factor)	嗜中性细胞分化因子	6
NP(Nucleoside phosphorylase)	核苷磷酸化酶	14
Npa(Nail－Patella syndrome)	甲－膑综合征	9
OPCA(Olivopontocerebellar atrophy Ⅰ)	橄榄体桥脑小脑萎缩Ⅰ型	6
P(Pblood group)	P 血型	6
PA(Paget disease of bone)	Paget 骨病	6
PepA(Peptidase A)	肽酶 A	18
PepB(Peptidase B)	肽酶 B	12
PepC(Peptidase C)	肽酶 C	1
PepA(Peptidase D)	肽酶 D	19
PepS(Peptidase S)	肽酶 S	4
Pg(Pepsinogen)	胃蛋白酶原	6
PGK(Phosphoglycerate Kinase)	磷酸甘油酸激酶	X
PGM－1(Phosphoglucomutase－1)	磷酸葡萄糖变位酶－1	1
PGM－2(Phosphoglucomutase－2)	磷酸葡萄糖变位酶－2	4
PGM－3(Phosphoglucomutase－3)	磷酸葡萄糖变位酶－3	6

续表

6PGD(6 – phosphogluconate dehydrogenase)	6 – 磷酸葡萄糖脱氢酶	1
PRPPAT(Phosphoribosylpyrophosphate amidotransferase)	磷酸核糖焦磷酸酰胺基转移酶	4
PK – 3(Pyruvate kinase – 3)	丙酮酸激酶 – 3	15
PKU(Phenylketonuria)	苯丙酮尿症	1
PP(Inorganic pyrophosphatase)	无机焦磷酸化酶	10
PVS(Poliovirus sensitivity)	脊髓灰质炎病毒敏感性	19
RB – 1(Retinoblastoma – 1)	视网膜母细胞瘤	13
rC3b(receptor for C3b)	C3b 受体	6
rC3d(receptor for C3d)	C3d 受体	6
Rg(Rodgers blood group)(same as C4F)	Rodgers 血型	6
Rh(Rhesus blood group)	Rh 血型	1
RN5S(5s RNA genes)	5S RNA 基因	1
RP – 1(Retinitis pigmentosa – 1)	色素性视网膜炎 – 1	1
RRNA(Ribosomal RNA)	核糖体 RNA	13,14,15,21,22
RwS(Ragweed sensitivity)	豚草敏感性	6
SA6(Suiface antigen 6)	表面抗原 6	6
SA7(Suiface antigen 7)	表面抗原 7	7
SA11(Suiface antigen 11)	表面抗原 11	11
SA12(Suiface antigen 12)	表面抗原 12	12
SA17(Suiface antigen 17)	表面抗原 17	17
SAX(X – linked species (or surface)antigen)	X – 连锁特异抗原	X
Sc(Scianna blood group)	Scianna 血型	1
SHWT(Serine hydroxymethyltransferase)	丝氨酸羟甲基转移酶	12
SOD – M(Superoxide dismutase, mitochondrial)	线粒体超氧化酶歧化酶	6
SOD – S(Superoxide dismutase, soluble)	可溶性超氧化酶歧化酶	21
Sph – 1(Sherocytosis, Denver type)	Denver 型球形细胞增多症	8 或 12
SRY(Sex detemirnation rigion on Y chromasome)	性别决定基因	Y
SV40 – 7(SV40 – integration site – 7)	SV40 整合位点 – 7	7
SV40 – 8(SV40 – integration site – 8)	SV40 整合位点 – 8	8
SV40 – 17(SV40 – integration site – 17)	SV40 整合位点 – 17	17
TK – M(Thymidine kinase, mitochondrial)	线粒体的胸苷激酶	16
TK – S(Thymidine kinase, soluble)	可溶性的胸苷激酶	17
TPI – 1 &2(Triosephosphate isomerase – 1 and 2)	磷酸丙糖异构酶	12
TrpRS(Tryptophanyl – tRNA synthetase)	色氨酸 tRNA 合成酶	14
TsAF8(Temperature – sensitive(AF8) complement)	AF8 温度敏感性补偿作用	3
UGPP – 1(Uridyl diphosphate glucose Pyrophosphorylase – 1)	尿苷二磷酸葡萄糖焦磷酸化酶 – 1	1
UGPP – 2(Uridyl diphosphate glucose Pyrophosphorylase – 2)	尿苷二磷酸葡萄糖焦磷酸化酶 – 2	2
UMPK(Uridine monophosphate kinase)	尿嘧啶单磷酸激酶	1
UP(Uridine phosphorylase)	尿嘧啶磷酸化酶	7
WS – 1(Waardenburg syndrome – 1)	Waardenburg 综合征	9
(W – AGR)(Wilms tumor – aniridia/ambiguous Genitalia/mentalretardation)	Wilms 瘤、无虹膜、精原细胞瘤、发育迟缓	11
XP – E(Xeroderma pigmentosum, Egyptian)	埃及着色性干皮病	9

图 10-26 人类中期染色体 C 带核型。(转引自 B.Lewin:Genes,Ⅳ 1990,Fig.26-18)

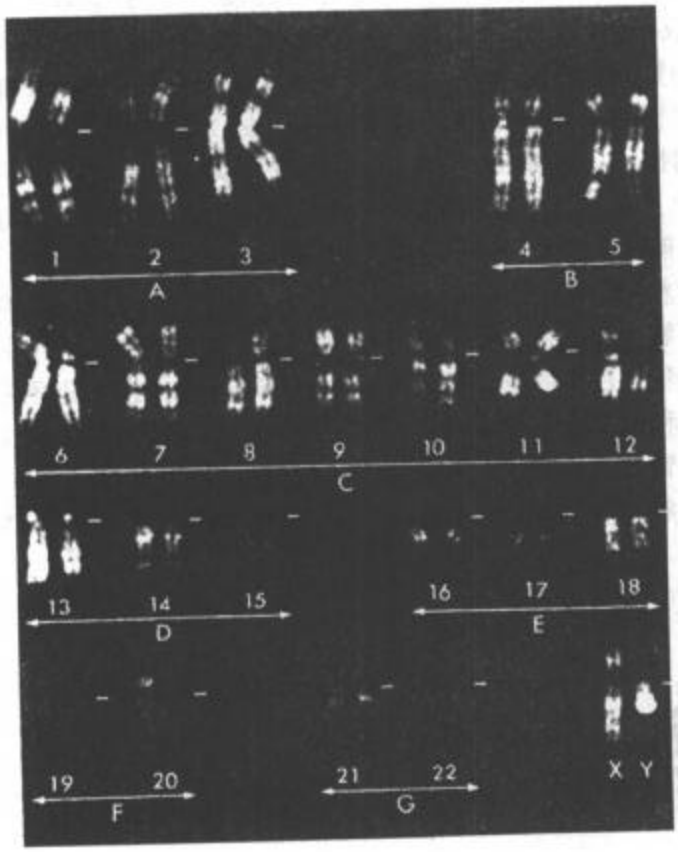


图 10-27 人类男性中期染色体 Q 带核型。(引自 Russell,P.J.:Genetics,3rd ed.,1992,Fig.10-14)

期保留。

R 带（R－Bands）是和 G 带相反（reverse）的带，即 G 带的深染区正是 R 带的浅染区，G 带的浅染区又正是 R 带的深染区。R 带也就是反带的意思。这种方法是由 Dutrillaux 和 Lejeune（1971 年）建立的。方法是将中期未染色的片子放在 pH 4～4.5。温度为 88℃的 1mmol/L 的 NaH_2PO_4 溶液中温育，然后再染色即可染成 R 带。

下面我们举一个关于大熊猫分类起源的问题为例来说明分带技术的应用。大熊猫（*Ailuro poda melanoleuca*）是我国特产的珍稀动物，从进化的观点来看，它是最神秘的动物之一。1869 年到 20 世纪 80 年代末，生物学家对于大熊猫的分类和起源一直存在着争论，有三种不同的观点。1869 年当时在中国的一位法国传教士和博物学家 Pere Armand David 首先对大熊猫进行了描述并定为一种新的物种，次年巴黎自然博物馆的馆长 Alphouse Milhe-Edwards 研究大熊猫的骨骼，断定大熊猫与同产于中国的小熊猫（*Ailurus julgens*）之间的亲缘关系要近于和熊的，确信大熊猫很像熊和浣熊共同祖先的后裔。此后关于大熊猫的起源问题有 40 余篇论文发表，但仍是众说纷纭，莫衷一是。1964 年芝加哥历史露天博物馆哺乳动物部主任 D.Dwight Davis 发表了一篇关于大熊猫解剖的详细论文，并根据 50 个器官的比较研究结果认为大熊猫是熊科的一员。但不久英国的 Desmond Morris 和 R.F.Ewer 根据其举止和形态分别断言大熊猫属于浣熊科。到 20 世纪 80 年代美国的动物学家 John F.Eisonbeg 和 George B.Schaller 建议把大熊猫单列一科，即大熊猫科，这一观点在 1986 年也取得我国一些动物学家的赞同。到底大熊猫的起源如何呢？这个珍贵又神秘的动物样子看起来像熊，但很多特征又不像熊，和小熊猫相似都是食草动物，且不须冬眠。它特殊之处是有相对拇指，它的叫声既不是嗥叫，也不是像羊一样发出咩咩的叫声。其染色体核型在数量上（22 对）排列上与熊的染色体（37 对）相比要更类似与小熊猫的染色体（22 对），而且熊属中的熊 37 对染色体都属于端着丝粒，而大熊猫的 22 染色体其中多数是非端着丝粒的，似乎很明确地表明熊和熊猫不属于同一科。这些事实使生物学家们感到困惑不解。但进一步地做 G 分带染色，结果使人豁然开朗，原来 6 种熊的带型实际上和大熊猫染色体带型是非常相似的，熊猫的中央着丝粒染色体（1，2，3）分别是熊的 2，3；1，9 和 6，16 染色体的带型相同（图 10－28）。意味着大熊猫可能由于熊科的祖先染色体长臂易位融合等进化而来，也就是说大熊猫起源于熊科，和小熊猫的亲缘关系要远于棕熊。现已在熊科下面单列一个新的大熊猫亚科，这个结果又得到了分子

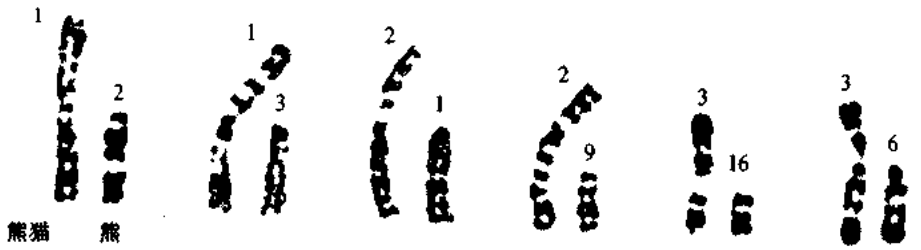


图 10－28　大熊猫（*Ailuropoda melanoleuca*）染色体 G 带型和棕熊的染色体 G 带型的比较。表明：大熊猫的 1 号染色的一条臂和棕熊的 2 号染色体同源，另一条臂和棕熊的 3 号染色体同源。大熊猫的 2 号染色体与棕能的 1 号和 9 号染色体同源。大熊猫的 3 号染色体与棕熊 6 号和 16 号染色体同源。

（引自 Stephen J.O'Brien：科学（中译本），1988，3，p35－40）

进化方面的许多证据，从而终于解决了100多年悬而未决的问题。

第五节 单一序列和重复序列

真核基因组是十分复杂的，不像原核生物多由单一序列组成，而是由各种重复序列和单一序列组成。

一、单一序列

单一序列(unique sequence)又称非重复序列，是指在基因组中只有1个或几个拷贝，真核中占40%～70%。真核基因组中大多数基因是单拷贝的，如卵清蛋白，丝心蛋白等基因。但单一序列在真核生物中大部分不编码，编码的序列仅占百分之几。原核生物除了短片段的回文序列(反向重复)和18S、28S、5S rRNA和tRNA基因以外皆为单一序列，这些单一序列基本上都编码。真核生物中单一序列编码的基因在表达时常有两步放大作用。如蚕的丝心蛋白基因只有一个拷贝但可转录产生10^4mRNA，而每个mRNA又可反复使用产生10^5个丝心蛋白分子。而重复序列编码的基因就没有第二步放大作用。

二、重复序列

短片段的重复序列按不同的排列方式可分为三种类型：①**正向重复**(direct repeats)又叫顺向重复，这种重复序列的方向是相同的。比如λ噬菌体的复制起点中的重复序列(见第12章)以及真核转座因子Ty，copia的两端(见23章)，插入序列，转座子的两端的宿主序列(见23章)以及端粒结构等；②**反向重复**(inverted repeats)的方向完全相反，呈两侧对称的序列，常存在于插入序列和转座子两端的结构元件中；③回文序列(Palindromic sequence)是一种旋转对称，如$\begin{matrix}5'\text{-GTGAGCTCAC-}3'\\3'\text{-CACTCGAGTG-}5'\end{matrix}$，其特点是正读、反读意思都一样，像中文中文字游戏的对联："画上荷花和尚画，书临汉字翰林书"一样。回文序列常存在各种蛋白质结合位点，调控序列中如操纵子，衰减子等。限制酶切位点也常常是由6～8个碱基组成的回文序列。

长片段的重复序列分为轻度重复、中度重复和高度重复三种。

三、轻度重复序列和中度重复序列

轻度重复序列是指在基因组中含有2～10拷贝的序列，如酵母tRNA基因、人和小鼠的珠蛋白基因等。

中度重复序列是指长约300bp在基因组中约有10到几千个拷贝的序列如rRNA和tRNA基因(表10-5)。tRNA基因一般都分布于基因组中，而rRNA常集中分布于核仁形成区。(图10-29)。不同的rRNA的基因情况也不同。在爪蟾中，5S rRNA有两种类型，一种是卵母细胞型，有24 000拷贝，存在于所有细胞中集中分布，但仅在卵母细胞中才表达。另一种是体细胞型，有400拷贝，分散分布于基因组中，在体细胞和卵细胞中都能转录表达，但在卵细胞中表达较弱。这样一组相同的或相关的基因排列在

表 10－5　几种生物基因组中多拷贝 rRNA、tRNA 基因

	18S/28S基因	5S基因	tRNA基因
大肠杆菌	7	7	60
啤酒酵母	140	140	250
黑腹果蝇	250(X)/150(Y)	165	850
人	280	2000	1300
爪蟾	450	24000	1150

B. Lewin: Genes Ⅴ, 1994, Table 24.5

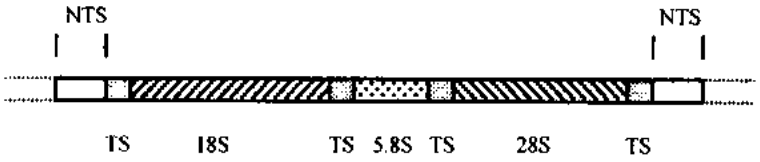


图 10－29　非洲爪蟾 rRNA 基因的串联重复排列。TS：可转录的间隔(spacer)序列；NTS：不可转录的间隔序列

一起称为**基因簇**(gene cluster)。另一个相关的名词是**基因家族**(gene family)，其含义和基因簇是不同的。基因家族是指由一个祖先基因经重复和变异(见第 25 章)所产生的一组同源基因，称基因家族。不同的基因家族其成员的多少(表 10－6)结构和功能的相似性都不相同，分散的基因家族其成员不在同一基因簇内，至少不在同一染色体上，串联基因家族的成员都集中串联在一起。有的基因家族成员具有和原有基因相同的结构和功能，如小鼠珠蛋基因家族有 5 个基因，一个是胚胎型的(α-*emx*)，另两个是成体型的(α-*ad1* 和 α-*ad2*)，它们都在第 11 号染色体上，这三个基因都可以编码，α-*ad1* 和 α-*ad2* 所编码的蛋白质仅在第 68 位上的氨基酸不同。还有两个成员是没有功能的假基因(α-ψ3 和 α-ψ4)分别位于第 15 和 17 号染色体上。表明同一家族成员也有的并不集中在一起，甚至在别的染色体上，同一基因家族的成员也有的是没有功能的假基因，人类的血红蛋白基因家族也是如此。**假基因**(pseudogene)是基因家族中因突变而失去功能的基因，不能产生具有生物活性的蛋白。和假基因相关的是**加工假基因**(processed pseudogenes)。假基因的结构上有以下的特点：①缺少正常的内含子；②3′末端有多聚腺苷酸；③5′端的结构和 mRNA 的 5′端十分相似；④两侧有顺向重复序列的存在(见第 25 章)，因此人们推测它似乎和 mRNA 一样经过了转录后加工，因此也称其为加工假基因；也有人认为其两侧有正向重复，像是反转录转座子的产物(见第 25 章)，所以又称其为**反转录基因**(retro gene)。这种假基因是因为没有启动子而不能表达。另一组没有活性

表 10－6　不同基因家族的组成

基因	基因家族成员数
肌动蛋白	5～30
角蛋白	＞20
肌球蛋白	5～10
微管蛋白	3～15
昆虫卵壳蛋白	50
珠蛋白	达到5
免疫球蛋白可变区	500
卵清蛋白	3
组蛋白	100～1000

的基因是由于丢失了原基因 5′或 3′区域所致。

另一些基因家族其成员的结构发生了较大的变化，可是仍执行相同的功能，例如人类的珠蛋白的 α 链和 β 链基因簇(图 10－30)。还有的基因家族成员，其结构和功能都发生了较大的变化，如编码肌红蛋白和血红蛋白的基因。复杂的多基因家族中各个成员都不相同，转录的方向也不同。如海胆和果蝇的组蛋白基因家族，每个家族 5 个基因串联在一起，作为一个单位，重复多达千次(图 10－30)。脊椎动物 β－珠蛋白基因簇

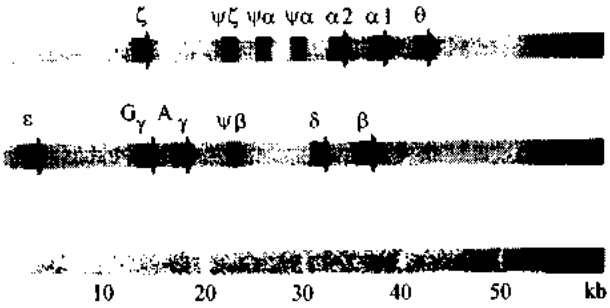


图 10－30 人类血红蛋白的 α 和 β 基因簇。(引自 B. Lewin: Genes, Ⅵ 1997, Fig 5－8)

的成员不仅结构和功能不同，表达的时间也不同，有的在早期胚胎发育时期表达，有的在胚胎晚期表达，有的要到发育成成体才表达。由串联重复序列中派生出来的单个基因，远离其家族，表达也会发生改变，这样的基因像是远离家庭的游子，故名为**孤散子**(orphons)。

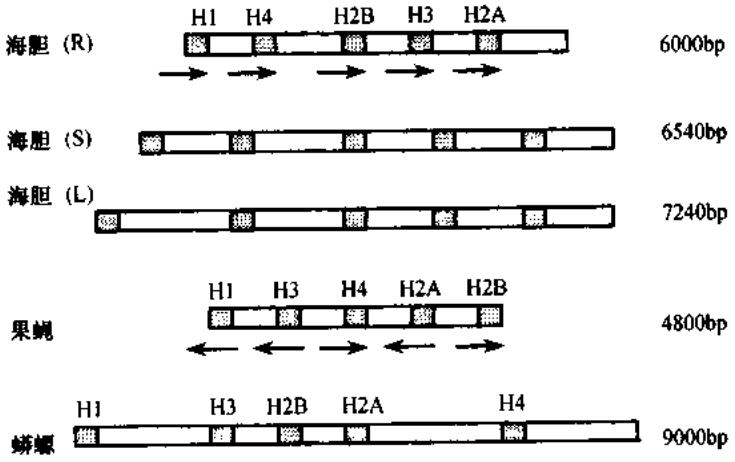


图 10－31 组蛋白基因簇的重复单位。图例：■ 基因；□间隔区；→转录方向。(仿 B. Lewin: Genes, Ⅵ 1990, Fig 26－10)

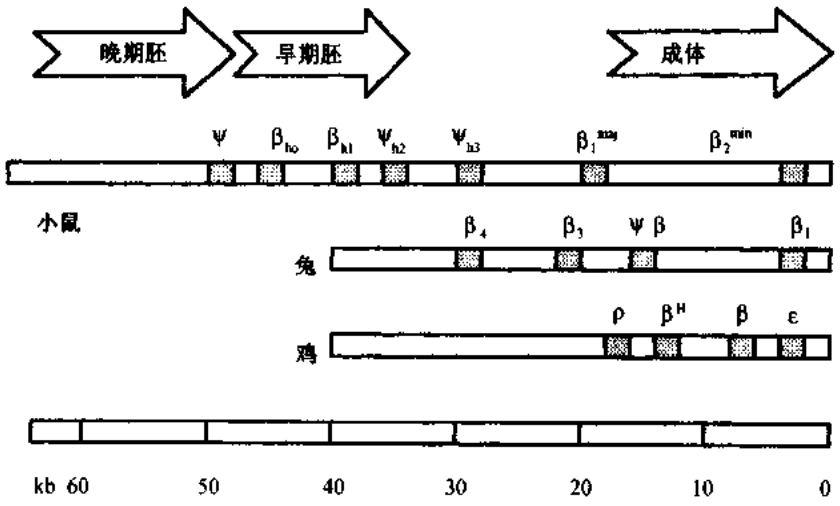


图 10－32　脊椎动物中的 β－珠蛋白基因簇和假基因。（仿 B. Lewin: Genes, Ⅵ 1997, Fig. 23－3）

四、*Alu* 家族

人类的 *Alu* 家族和小鼠 B 序列家族是十分典型的分散基因家族。在人类的基因组中有一种中等重复序列，长约 300bp，30 万个成员分散分布在单倍体基因组中，在其 170bp 处有一个限制性酶 *Alu* Ⅰ的酶切位点，故称这个重复序列为 *Alu* 基因家族（Alu family），大约平均每隔 6kb 左右就有一个 *Alu* 序列，因此它可能含在内含子或基因附近的序列中。所以它可作为人类 DNA 片段的特异标记。

Alu 家族的每个成员彼此都很相似，由 130bp 的串联重复序列而组成，在二聚体的右半个中部有 31bp 无关的序列插入在里面。这个插入序列（IS）来自 7SL RNA（是信号识别蛋白 SRP 的一个成分）。7SL RNA 长 300nt，其 5′端的 90nt 和 *Alu* 序列左端同源，其中央的 160 个碱基和 *Alu* 并不同源，而 3′端的 40 个碱基和 *Alu* 右端同源（图 10－33）。

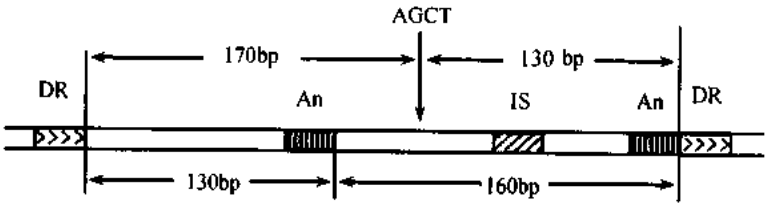


图 10－33　*Alu* 序列的基本结构。DR：顺向重复；An：多聚腺苷；IS：插入序列

Alu 家族的成员和转座子相似，两端有短的正向重复序列存在了。然而十分奇怪的是 *Alu* 家族的每个重复片段的长度不同。因为它们可能有 RNA　pol Ⅲ转录而来，所以每个成员可能带有下游启动子。

Alu 家族的广泛存在意味着它具有某种功能。只是目前尚未能搞清楚。部分 *Alu* 序列中有 14bp 区域和乳头瘤病毒、乙型肝炎病毒的复制起始区有同源性，因此人们推测 *Alu* 家族可能和真核基因组的复制区相连接。但与此相矛盾的是 *Alu* 序列的成员数要比推测的复制区多 10 倍。

Alu 家族的某些成员能在体外由 RNA pol Ⅲ转录成 SnRNA，在中国仓鼠中，*Alu* 家族成员可以转录。其转录本常在其他转录单位的边上。

在小鼠中有和 *Alu* 序列同源的序列称为 B1 基因家族，约有 5 万个成员。但在中国仓鼠中仍称 *Alu* 序列。B1 的重复单位长 130bp，相当于 *Alu* 序列的一个单体，和 Alu 序列的同源性约 70% ~ 80%。 *Alu* 序列也称为短的分散因子(short interspersed elements, SINEs)。它是由 RNA 聚合酶Ⅲ转录的。在哺乳动物中还有另一种长的分散因子 (long interspersed elements, LINEs)，是由 RNA 聚合酶Ⅱ转录的。如 L1 长 6500bp 左右，在基因组中约 6 万个拷贝。属于一种转座因子 (见第 23 章)。在每一种哺乳动物中只有单个 LINEs 家族。在不同物种中的 LINEs 家族之间具有较高的同源性，至少在一个家族中有的 LINEs 是编码蛋白质的，但这种蛋白尚未描述过。

所有真核生物中都具有 SINEs 和 LINEs，但有不同的比例，例如果蝇和鸟类含 LINEs 较多，而在人类和蛙中则含 SINEs 较多。

五、高度重复序列

高度重复序列是一种简单的重复序列，有的重复单位长度不超过 6bp，但在基因组中重复可达十万次。由于他们在很低的 C_0t 值时变性，所以易于分离。在小鼠中这些序列(占基因组的 8%)含 G - C 为 30%，远比基因组中其他部分要低得多，和基因组中的主体 DNA(含 G - C 为 42%)是不同的(占基因组的 98%)。当将 DNA 切成片段进行氯化铯密度梯度超离心时，由于富含 A - T 片段的浮力密度小，在离心管中常常单独形成一条较窄的带，在主体 DNA 带的上面，故称为卫星 DNA(satellite DNA)。卫星 DNA 存在于很多真核基因组中。他们可能比主带轻，也可能比主带重。

通常基因组中的高度重复序列是可以从卫星 DNA 中分离出来。当一个高度重复序列并未分离成一个卫星带时，我们常常可以通过其他的方法，如复性动力学等方法将这种“隐藏”在基因组中的高度重复序列分离出来。它们同样是高度重复的，但 G - C 的含量和基因组中其他 DNA 没有明显的差别，所以用密度梯度离心就分不出一条卫星带来。这种高度重复序列我们就称它为**隐蔽卫星 DNA**(cryptic satellite)。无论是隐蔽卫星还是明显卫星结构特点都相同，属于不同类型的高度重复序列。在一个基因组中常含有一种类型以上的卫星 DNA。

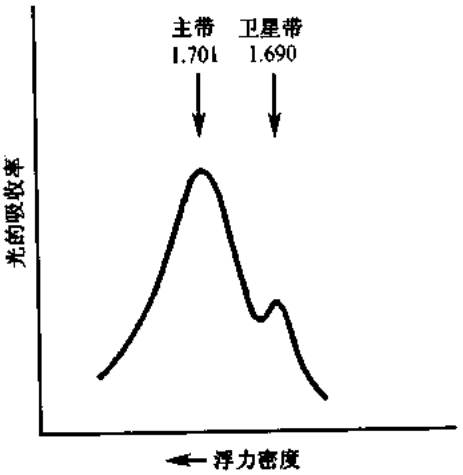


图 10 - 34 小鼠 DNA 经 CsCl 密度梯度离心显示出主带和卫星 DNA 带。

人们认为高度重复序列并没有什么功能，但在长期的进化中不仅没有被淘汰而且还以“高度重复”的形式存在，未免过于“自私”，因此也称其为“自私 DNA(selfish DNA)”。其实高度重复序列不见得没有任何功能，也许目前尚未发现而已。1982 年 Singh 等分离了雌蛇的高度重复序列，以此为探针进行分子杂交，在小鼠的 Y 染色体上找到了同源区域，分析结果显示这段可能和性别决定有关 (但在人类 Y 染色体上未发现这段序列)，因此又称其为伊甸园 DNA

(Garden of Eden DNA)。

还有一种重复序列称为"小卫星"(minisatellite)或同向重复序列可变数(variable number tandem repeat, VNTRs)区。这是一种特殊的串联重复,在不同个体和基因组的不同位点上数目都不同。在人类中 VNTRs 位点是 1 ~ 5kb 的序列,此序列由长 15 ~ 100nt 的重复单位组成。当将人类的总 DNA 提出后,用限制性内切酶切成不同长度的片断(各种 VNTRs 上都没有酶切位点),然后以 VNTRs 中的特异序列为探针进行 Southern 杂交,可发现阳性片断的长度各不相同。由于不同个体的这种串联重复的数目和位置都不相同,所以 VNTRs 的 Southern 杂交带谱就具有高度的个体特异性,人们就称其为 DNA 指纹(DNA fingerprints)(图 10 - 35)。可用于亲子鉴定、法医鉴定等。血液和精液都可以用作标本,甚至极少量的标本,比如连在一根头发上的毛囊细胞,其 DNA 经 PCR 扩增后也可用来检测。

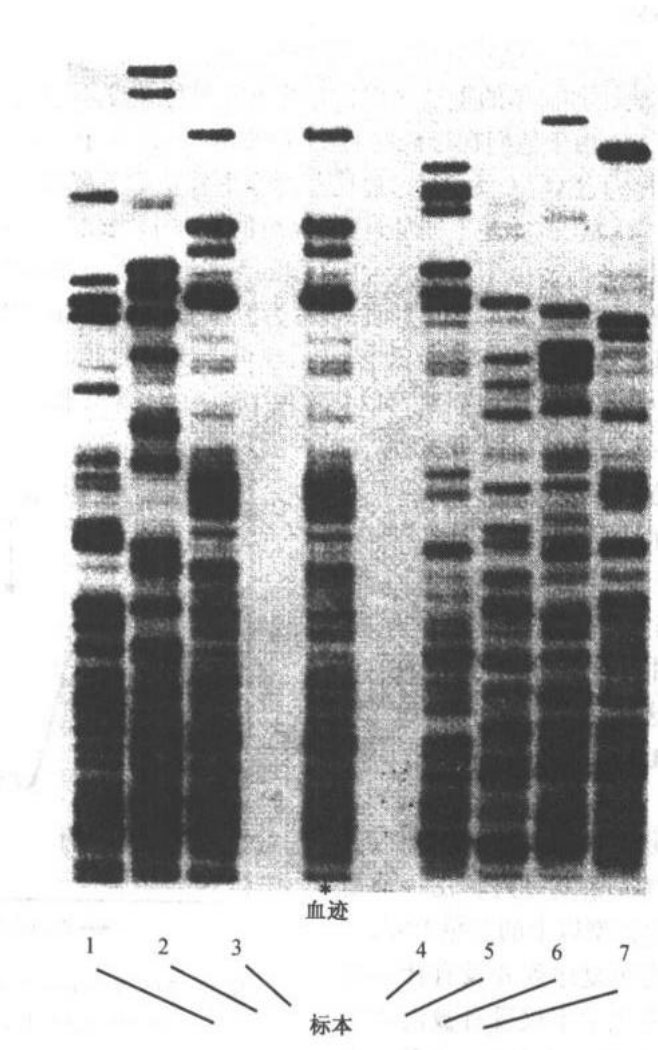


图 10 - 35　人类 DNA 指纹。★表示犯罪嫌疑人的血迹和各种来源血液的指纹比较,其中标本 3 和犯罪嫌疑人的 DNA 指纹相同。(转引自 Griffiths A. J. F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993. Fig. 6 - 25b)

人类 VNTRs 的第一个探针是在 1985 年由 Alec Jeffries 在肌红蛋白基因的第一个内含子中分离出的 132bp 的重复序列，内含 4 个重复单位，每个单位长 33bp，内有 13bp 为核心序列（GGAGGTGGGCAGG），他就用含有核心片断的序列为探针。尽管各串联重复序列并不相同，但都存在核心序列。每个串联重复序列含有 5～50 重复单位，普遍存在于哺乳动物中，在群体中 VNTRs 存在着多态性。因此和限制性片断长度多态性一样，也有可能作为等位基因的一种标记，即 A 和 a 一对等位基因所连锁的 VNTRs 可能是不同的，因此可以通过鉴别不同的 VNTRs 作为标记来确定 A 和 a。

VNTRs 多变的原因是个体的等位基因可能连锁着不同数量的重复单位，在减数分裂时它们之间可以发生重组，产生新的组合。

在VNTRs 序列中发生交换的频率是很高的，每 kb DNA 约为 10^{-4}。在人类基因组中核心序列为 10～15bp 的富含 G·C 的序列，而且 G 和 C 分别集中于两条链上，如 GGGCAGGA × GCCCGTCCT × C。每一个单个的"小卫星"都有一个核心序列的变异体，但用一个用核心序列制备的探针来做 Southern 杂交可检查出 100 个小卫星序列来。

我们将以上的概念联系起来画出真核染色体中各种不同的序列，使我们对真核染色体的复杂性留下一种具体的印象（图 10－36）。

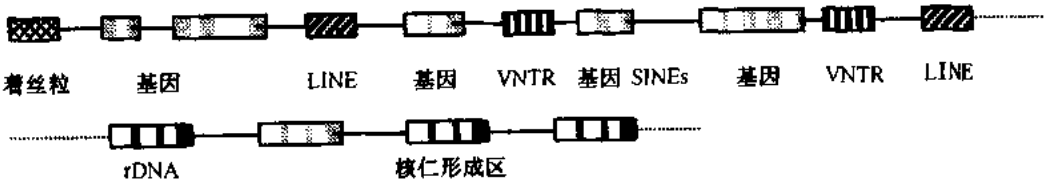


图 10－36(a)　真核生物染色体上的各种重复序列和结构元件。（仿 Griffiths A.J.F.et al:An Introduction to Genetic Analysis,5th ed.,1993.Fig16－28）

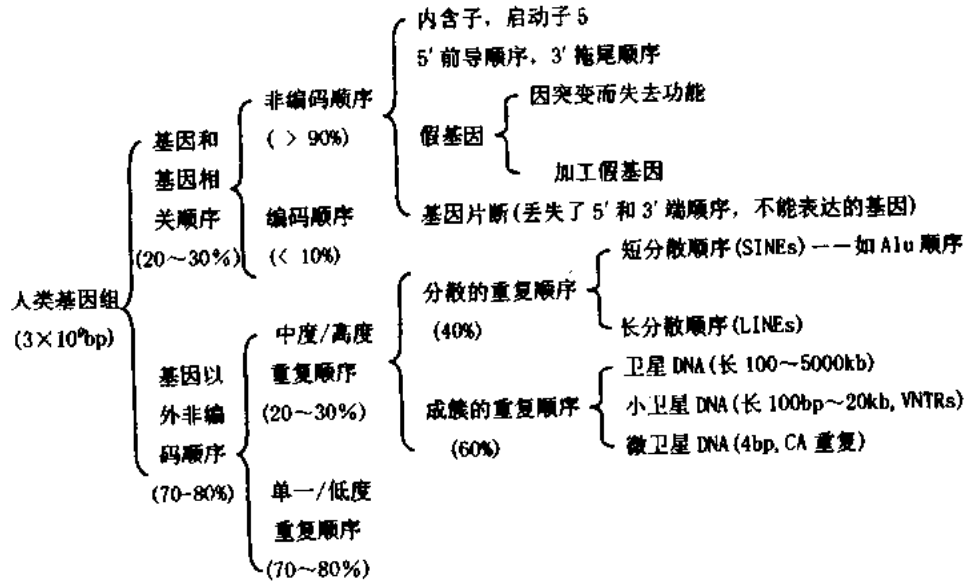


图 10－36(b)　人类基因组的序列。

（参考 P.C.Winter *et al*:Instant Notes in Genetics,BIOS Scientific Publihers Limited,1998,Fig 1）

第六节　C 值矛盾

在每一种生物中其单倍体基因组的 DNA 总量是特异的，被称为 **C 值**(C-value)。DNA 的长度是根据碱基对的多少推算出来的。各门生物存在着一个 C 值范围(图 10－37)，在每一门中随着生物复杂性的增加，其基因组大小的最低程度也随之增加。

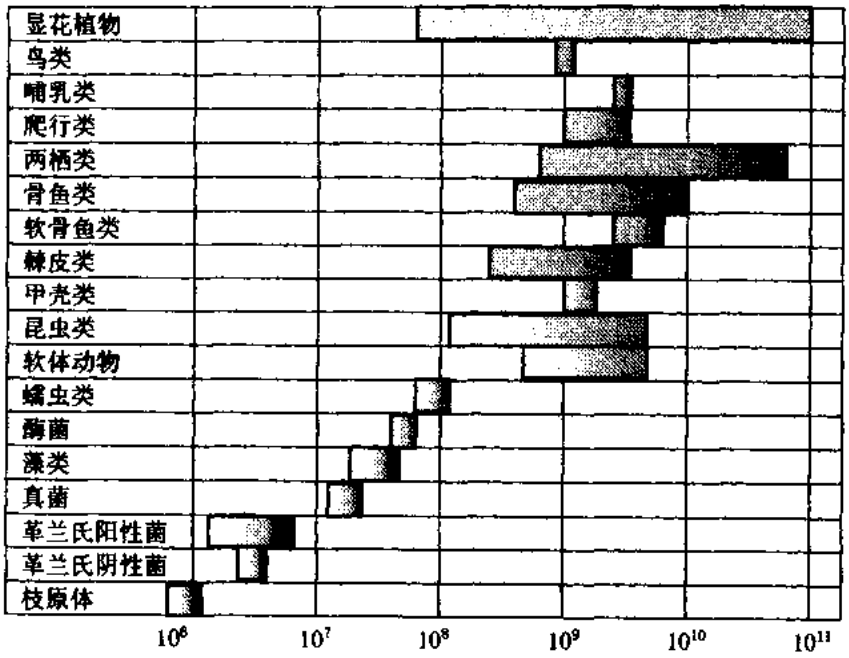


图 10－37　不同门类生物的 C 值分布(仿 B. Lewin: Genes, Ⅵ 1997, Fig. 21－1)

C 值的资料表明，在不同的门类中 C 值的变化是很大的。相对比较简单的单细胞真核生物像啤酒酵母，其基因组就有 1.75×10^7bp 大约是细菌(如 *E. coli*)基因组的 3～4 倍。最简单的多细胞生物秀丽隐杆线虫其基因组有 8×10^7bp，大约是酵母的 4 倍。看来生物的复杂性和其 DNA 含量之间有较好的相关性。但我们可看到在其他的一些门中，这种相关性有的并不存在。实际上一个门中的 C 值变化并没有一定的规律。例如在哺乳类、鸟类和爬行类的 C 值变化范围都很小，而在两栖类中这种变化范围增大，而植物的 C 值变化范围更为宽广，常成倍成倍地增加。C 值和生物结构或组成的复杂性不一致的现象称为 **C 值矛盾** (C-value paradox)。现在我们还不能完全解释这种矛盾。在一定意义上说，生物类群中 C 值变化范围宽就意味着在某些生物中有些 DNA 是冗余的，不能编码有功能的活性物质。DNA 总量变化范围的产生至少有一个原因，即在染色体上存在着不同数目的重复序列，这些重复序列是不表达的。

第七节　断裂基因

真核生物的基因组十分复杂，DNA 的含量也比原核生物的大得多。前面已介绍几种噬菌体由于基因组很小，但又要编码一些必不可少的蛋白，碱基显然不够用，这样

不仅几乎所有的碱基都参加编码，而且在进化中还出现了“重叠基因”，以有限的基因编码更多的遗传信息。真核基因组正好相反，DNA 十分富余，这样不仅无需“重叠基因”，而且很多序列不编码，如重复序列、**间隔序列** (spacer) 和**间插序列** (intervening sequence) 即内含子(intron)等。但不编码并不等于没有功能。有的我们可能还不了解，如重复序列。间隔区和间插序列这两个概念是不同的，间隔区是指基因间不编码的部分，有的转录称转录间隔区(TS)，有的不转录称为非转录间隔区(NTS)。间插序列是指基因内部不编码的区域，也称内含子，在初始转录本中存在此序列，但在加工后将被切除掉，所以常不作为翻译的信息。间隔区常常含有转录的启动子和其他上游调节序列(见第 12 章)。有的内含子也可以编码，如成熟酶和内切酶等。

一、内含子的发现

1977 年美国的 Sharp 和 Roberts 两组科学家分别同时发现了**断裂基因** (split gene)，Crick 称此为分子遗传学上的一次微型革命，这项发现与 1993 年荣获了诺贝尔奖。

就在 Sharp 和 Roberts 发现内含子之前，法国的科学家 Chambon 也率领一个小组进行了有关实验。他们在思考一个问题，鸡的输卵管分泌卵清蛋白、卵粘蛋白和伴清蛋白，而其红细胞(鸟类的红细胞上有核的)只合成血红蛋白，那么两种组织之间 DNA 有什么不同呢？于是他们提取两种组织的 DNA，分别用酶(*Eco*R Ⅰ和 *Hin*d Ⅲ)切成几段，走电泳，再用卵清蛋白 mRNA 来制备 cDNA 探针和以上两种片段进行 Southern 杂交，结果两种组织中的 DNA 不论用哪一种酶来切，都出现了相同的多条阳性杂交带。图 10－38 表明不论是红细胞还是输卵管的细胞，虽基因表达不同，但 DNA 的结构没有差异。但令它们不解的是卵清蛋白 cDNA 序列内并没有 *Eco*R Ⅰ和 *Hin*d Ⅲ的切点，为什么会出现多条阳性带。

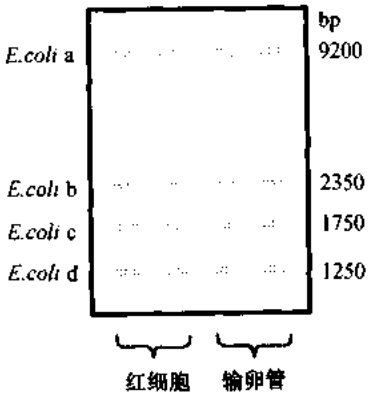


图 10－38 鸡卵清蛋白 cDNA 探针和红细胞及输卵管 DNA Southern 杂交的结果。

几个月以后 *Berget*，*Sharp* 小组和 *Roberts* 小组同时发现了腺病毒外壳蛋白六聚体基因(*Hexon gene*) 前导区有断裂现象(图 *10－39*)。他们用限制性内切酶 *Eco*R Ⅰ和 *Hin*d Ⅲ分别消解腺病毒的 DNA，得到了大小不同的很多片段，分别选择两种酶切片段中的最大的 A 片段 DNA 和 Hexon 的 mRNA 进行杂交，在电镜下可以观察到 *Eco*R Ⅰ酶切的 A 片段的 3′段可以和上述 mRNA 形成杂合双链，但在杂合双链的 5′端逸出 3 个单链 DNA 环，说明它们不能和 mDNA 完全互补。他们的解释是 Hexon 基因 5′端出现了断裂，即有些区域在 mRNA 中已被删除了，所以出现了不配对的单链环，而 *Hin*d Ⅲ酶切的 A 片段是 Hexon 的 3′段区域，不含有这一断裂区域，故不出现单链环。

Berget 在冷泉港作了有关发现断裂基因的学术报告，提出在 Hexon 基因内近 5′端有不编码的部分。Chambon 听了报告后便意识到，他们的实验结果也是可以用断裂基因来解释的。即卵清蛋白的基因上可能有多个断裂区(内含子)。在这些断裂区上有酶切位点的存在，可将卵清蛋白基因切成大小不同的片段，但它们都可以和 mRNA 进行

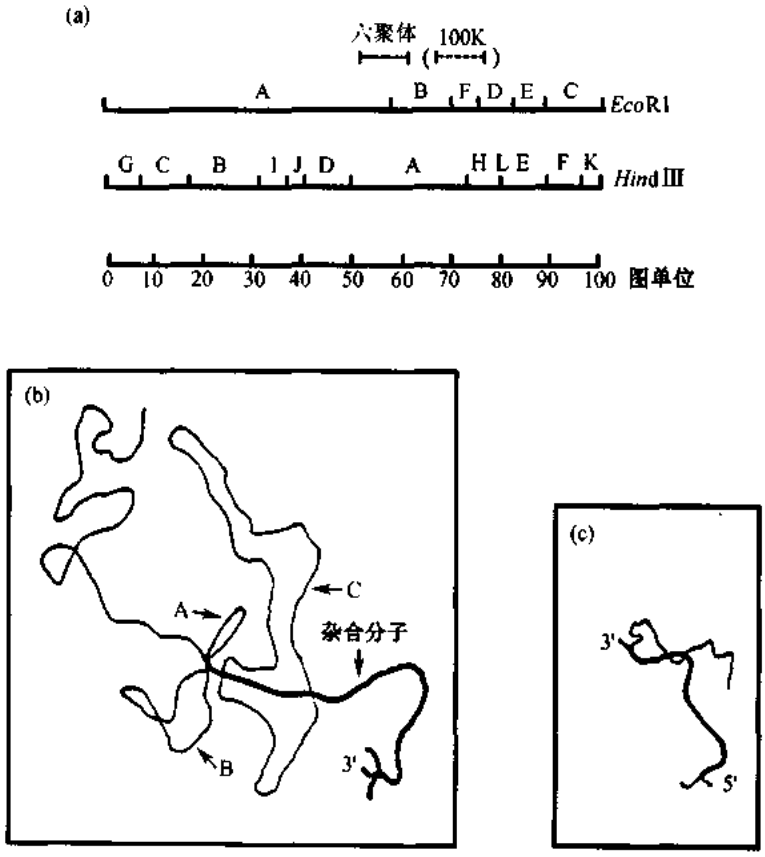


图 10－39　腺病毒六聚体蛋白基因限制性酶切片段 A 和 mRNA 的分子杂交。(a) *Eco*R Ⅰ和 *Hind* Ⅲ的酶切图谱；(b)mRNA 和 *Eco*R Ⅰ酶切片段 A 杂交 3′端出现 3 个单链侧环；(c)mRNA 和 *Hind* Ⅲ酶切片段 A 杂交 5′端出现了双链区，但无侧环。(转引自盛祖嘉，沈仁权：《分子遗传学》，1988，复旦大学出版社，图 7－12)

杂交。事后 Chambon(1977，1981)等用 Berget 的实验方法进行了分子杂交，果然出现了 7 个单链 DNA 的环(图 10－40)，表面有个断裂区的存在，虽然错过良机，但证实了自己的推测仍然是很值得称赞的事。

1977 年继 Sharp 和 Roberts 后相继发现了兔子的 β 珠蛋白基因 (Jeffregs，A.J. 和 Flavell，R.A. 1977)、鼠的 β 珠蛋白基因 (Leder，P.1978) 和卵清蛋白基因 (Chamber 1977)等的断裂基因。

1978 年 Gilbert 创用了内含子 (intron) 和外显子 (exon) 两个名词，内含子是指在成熟的 mRNA 中不出现的序列 Exon 是指在成熟的 mRNA 中出现的编码序列。

二、内含子的结构特点

1. 内含子和外显子的分布

真核基因一般都含有内含子，也有少数基因不含内含子，如组蛋白基因，干扰素基因，酵母的多数蛋白质基因。1986 年 Chu F.K. 等发现 T4 噬菌体的胸苷合成酶基因

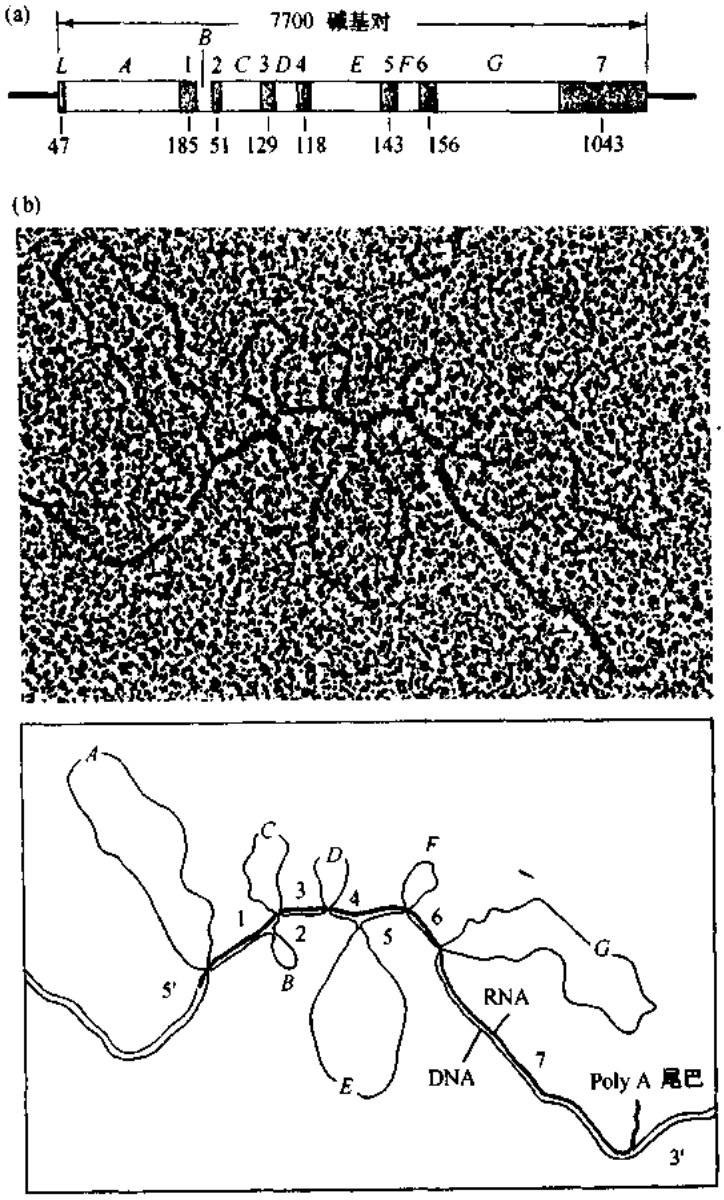


图 10－40 鸡卵清蛋白基因的结构。(a)外显子(L,1－7)和内含子(A－G)的排列;(b)经分子杂交 R 环部分从双链中逸出。(引自 Russell,P.J.:Genetics,3rd ed.,1992,Fig.13－8)

含有内含子。此外猿猴病毒 SV40 的 T 和 t 蛋白的基因中也含有长度不等的内含。不同的基因其内含子的大小多少是不相同的,有时悬殊很大。有的基因内含子少,如珠蛋白基因只有 2 个内含子,有的基因内含子很多,如鸡的胶原蛋白(1α2)基因有 50 个外显子。长的内含子如二氢叶酸还原酶,它有 5 个内含子,6 个外显子,mRNA 长 2kb,而 DNA 长超过 31kb、即 5 个内含子的总长度达到 29kb,比外显子长得多。

2. 内含子的相对性

内含子是相对的,一个基因的内含子可能是另一个基因的外显子。

如 SV40 病毒早期转录区编码了 T 和 t 两个功能不同的蛋白,这两个基因是重叠的,仅内含子不同。初始转录物的长度为 2530nt。大 T 蛋白成熟的 mRNA 中切除了 330nt

的内含子，余下的编码序列为 2200nt，翻译成大 T 蛋白的分子量为 100kDa。但小 t 蛋白内含子仅 30nt，位于大 T 内含子区域内，大 T 内含子其余部分（300nt）对于小 t 来说却是外显子。所以小 t 蛋白的 mRNA 长为 2500nt，由于在其内含子前有一个终止密码子，所以能翻译的部分并不长，小 t 蛋白的分子量也只有 18kDa（图 10－41）。

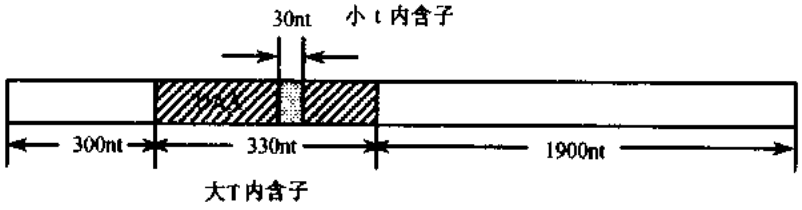


图 10－41　SV40 的 T 和 t 蛋白基因的内含子。

大鼠的肌钙蛋白 T（troponin T）也是一个很好的例子。它具有两种不同形式的蛋白，α 型和 β 型。都由同一个基因编码，这个基因共有 5 个外显子，其中 W，X 和 Z 是两种类型蛋白共有的外显子，另外两个 α 和 β 分别在不同类型的 mRNA 中表达，即 α 在 α 型中是外显子在 β 型中都是内含子，加工时被剪掉，β 外显子也是相似的情况，说明内含子是相对的（图 10－42）。

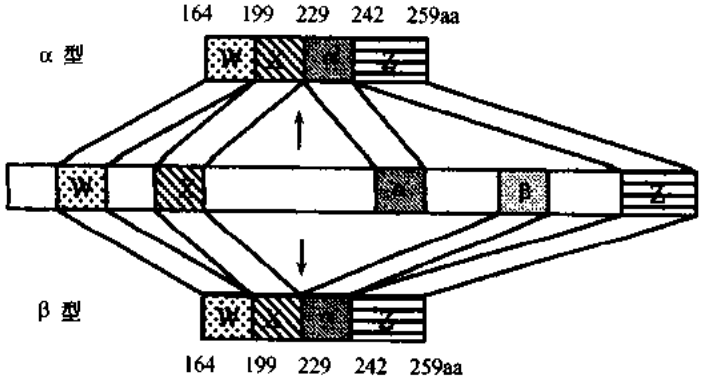


图 10－42　肌钙蛋白基因内含子的交替剪接，产生 α 和 β 两种类型的蛋白。
（仿 B. Lewin：Genes，Ⅳ，1990，Fig 25－7）

三、内含子的生物学意义

在真核生物的基因中普遍存在着内含子，既然它不编码，那么其生物学意义何在？在长期的进化过程中为什么它尚能保留安然无恙呢？Gilbert（1978 年）指出：断裂基因的存在可能有两种意义：①有利于储存较多的遗传信息，如 SV40 的 T 和 t 蛋白；②有利于变异和进化，若在交界序列处发生突变，就可能影响正常的剪切方式从而使蛋白的结构发生较大的变化。除此之外内含子还有以下的作用。

1. 有的内含子可编码内切酶

酵母的细胞色素 b 基因含有 3 个外显子（417bp，14bp 和 100bp）和两个内含子（765bp 和 1240bp），转录成 mRNA 后首先切除了内含子 1，形成未成熟的 mRNA。核糖体结合在这种 mRNA，利用外显子 1，2 和内含子 2 的部分序列（5′端 840nt）为模板，翻译成长为 423aa 的一种蛋白，叫成熟酶（maturase），此酶回过头来又可将未成熟的 mRNA 中的内含子 2 切除掉，产生成熟的 mRNA，最终以此 mRNA 为模板翻译成细胞色素 b（279aa）。细胞色素 b N 端的 144aa 和成熟酶 N 端的 144aa 是同源的（图 10－43）。

另一个 例子是酵母 ω 蛋白内含子的归巢（homing）。酵母中有两种不同的品系，ω^+

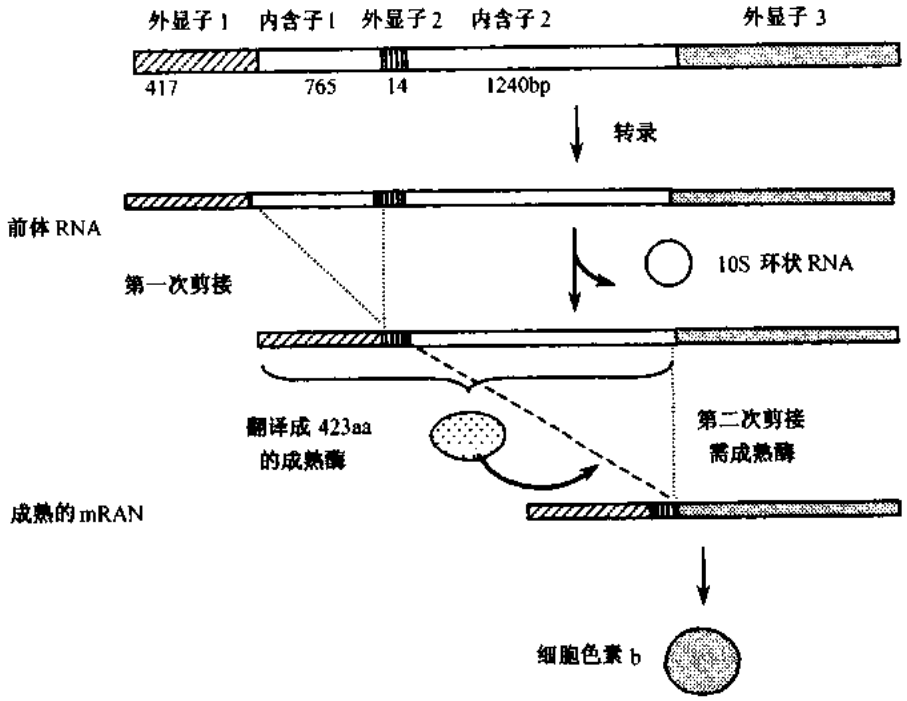


图 10－43　酵母细胞色素 b 基因的内含子 2 参与编码成熟酶，而成熟酶又反过来参与内含子 2 的剪接，产生成熟的 mRNA。

和 ω^-，ω^+ 有内含子，而 ω^- 在相应的位置上无内含子。若将 ω^+ 品系和 ω^- 进行杂交，按一般规律合子为 ω^+/ω^-，经减数分裂后应产生 2 个 ω^+ 孢子和 2 个 ω^- 的孢子。但实际结果是所有的孢子都是 ω^+。这是如何产生的呢？此是由于 ω^+ 品系的内含子可以编码一种内切酶，在靶位点将 ω^- 基因的 DNA 切开，而此内含子可以反转录成 DNA 双链，插入到切开的靶位点上，连接后使 ω^- 变成了 ω^+，这个过程就称为内含子的归巢（homing）（图 10－44）。

2. 调控

在内含子的相对性中我们举了 SV40 的 T 和 t 蛋白，以及大鼠肌钙蛋白的两种类型，从中可以看出真核基因组中虽然没有重叠基因，但通过内含子的相对性可使相同的一段 DNA 编码不同的蛋白，起到了调控的作用，并增加了遗传信息的储存。

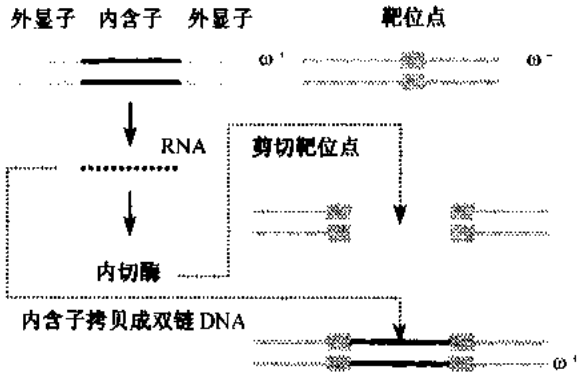


图 10－44　酵母 ω 蛋白内启子的归巢。(参考 B.lewin:Genes, Ⅵ, 1997, Fig.31－10)

3. 提高进化速率

蛋白质的三维结构与内含子和外显子的分布有密切的相关性。外显子常常对应于蛋白质的折叠区域(protein folding domains)或功能域(function domains)。如免疫球蛋白和β珠蛋白分子的各个结构域和外显子的分布是一致的。在进化中，外显子是稳定的，内含子的序列和长度是迅速随机突变的。重组易发生在外显子之间的内含子之中，这样可形成蛋白结构域的新组合，而减少在外显子上发生重组，若发生这样的重组可能导致蛋白质功能的破坏，内含子的存在提高了有效重组。在原核生物中，没有有性生殖的过程，不发生减数分裂，因此也不存在真核生物相互交换的同源重组，只发生部分外基因子和完整内基因子之间的单向重组。无论是转化还是接合，无论是转导还是转座都发生在编码区中，仅 λ 噬菌体的整合发生在基因间的 att 位点，这些重组往往会改变原核生物的性状。原核生物诸多机能都倾向一个字："变"。只有多变才能适应环境的改变，抵抗宿主的限制和抗生素的作用。因此无内含子是有利的。真核生物诸多机能都倾向于"稳"。因在长期的进化中，真核生物已产生了复杂的结构和各种系统适应了环境，新的变化常常对生物是不利的。内含子的存在也是有助于稳定蛋白结构的一种适应性。

由于内含子的存在，在此区发生的非法交换可以产生基因的重复，而进化中形成了基因族及同工酶的基因，在原核生物的基因组中没有内含子就不可能产生结构基因的重复，因此也没有基因家族的存在。

LDH 基因的进化也同样是通过重复和突变产生的。

由于内含子的存在在保留原有蛋白的同时，可产生一种新的蛋白，具有进化优势。由于在外显子和内含子的交界区发生突变，可产生新的 5′或 3′连接点，对于单个 3′位点来说可以和原来的 5′位点拼接，也可以和新的 5′位点结合，这样产生的不同的 mRNA 既可以保留原有的蛋白，也可以产生一种新的蛋白，而原核生物基因突变只有在编码区发生，这样产生了新的蛋白，但原有基因的产物就不可能形成了。

第十一章

DNA 的复制

基本概念

原核生物 DNA 复制是通过半保留机制进行的。DNA 双螺旋的两条链彼此分离，作为模板分别互补地合成两条新链。这种机制使得在细胞分裂时遗传信息得到精确地拷贝。

催化 DNA 合成的酶叫做 DNA 多聚酶或聚合酶。所有的 DNA 聚合酶都同样地催化 DNA 合成反应。利用 5′－三磷酸脱氧核苷为底物，延着 5′到 3′的方向合成新链。

DNA 聚合酶不能起始合成一条新的 DNA 链。所有新的 DNA 链都需要 RNA 引物，此引物是通过 DNA 引物酶等酶催化合成的，在不同的原核生物中合成引物的酶不尽相同。

在 *E. coli* 中 DNA 的复制需两种 DNA 聚合酶和其他的几种酶和蛋白。DNA 的合成在一条模板链上合成是连续的，而在另一条模板链上的合成是不连续的，这一过程称为半不连续复制（semidiscontinuous replication）。

真核生物 DNA 的复制是在细胞周期的 S 期进行的。它和原核生物的 DNA 复制过程基本相似。随着 DNA 的复制染色体本身也产生重复。当复制叉移动时，组蛋白组装成核小体。

DNA 复制的起动使细胞（原核和真核）进一步分裂。从这个观点出发，也可以说细胞的增殖取决于 DNA 复制是否起动。

在复制未完成时细胞是不会发生分裂的。因为只有复制完成才能触动细胞分裂，复制的基因组彼此分离，进入不同的子细胞，分离的单位是染色体。

第一节　半保留复制的验证

Watson 和 Crick 发表了 DNA 双螺旋模型之后不久，于同年又紧接着发表了 **DNA 半保留复制的复制机理**，这一建立在碱基互补基础上的机制为转录、修复、重组、分子杂交等奠定了基础。在他们提出半保留复制之前，普遍认为“蛋白质和核酸是彼此互补的，DNA 自我复制过程是通过核酸和蛋白质的交替合成”。但 Watson 和 Crick 认为：“DNA 自我复制并不依赖于特异的蛋白质的合成，DNA 双螺旋中的每一条互补 DNA 链，在合成它的一条互补新链时都可以作为模板”。复制时“连接互补链的氢键必须断裂，两条链还必须解缠和分离。很可能这种单链本身可以呈螺旋构型，并作为模板，游离的核苷酸（正确的多核苷酸前体）通过形成氢键自动地结合到它上面”。这就是 DNA 复制的**半保留模型**（semioconservetive model）。当时他们还不清楚复制的过程“是需要一种特异的酶呢？还是单螺旋能像一种酶那样有效地作用？”

当时人们对半保留模型是持怀疑态度的，因双螺旋模型中存在最大的问题是双链是如何打开的？其所需的能量从何而来？Watson 和 Crick 也承认这是“最大的问题”。人们为了避开打开双链这个难题，又提出了另外两个模型：一种是**全保留复制模型**（conservetive model），另一种是分散模型（Dispersive model）。在全保留模型中两条母链彼此结合，恢复原状，新合成的两条子链彼此互补结合形成一条新的双链。在分散模型中亲代双链被切成双链片段，而这些片段又可作为新合成双链片段的模板，新、

老双链片段又以某种方式聚集成"杂种链"(图 11－1)。这三种模型到底哪一种是准确的呢？人们用不同的方法雄辩地证实其中仅半保留模型是惟一正确的。现在我们首先介绍的 Meselson-Stahl 实验。

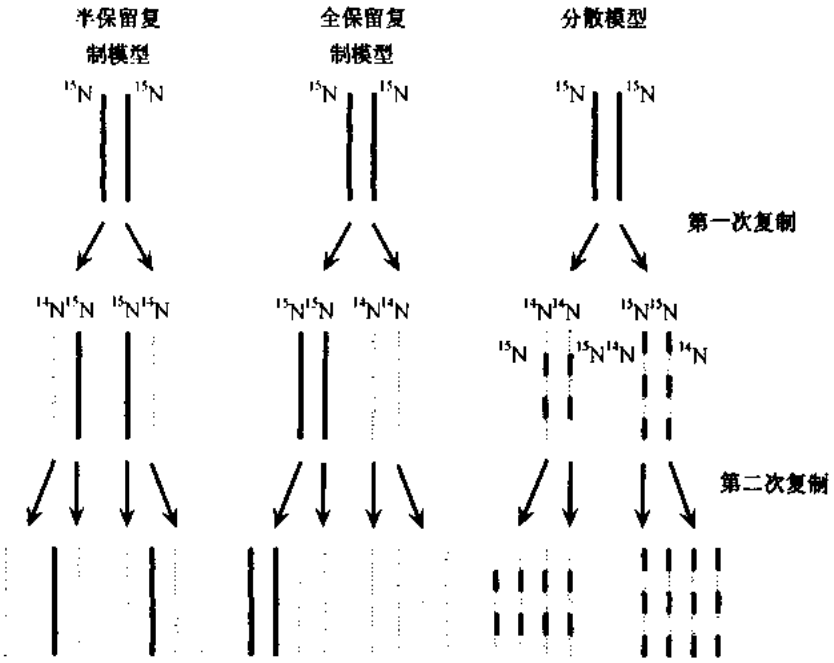


图 11－1　三种不同的复制模型。

一、Meselson-Stahl 实验

半保留复制模型于 1953 年就已经提出，直到 1958 年 Meselson-Stahl 才用实验证实其正确性，可见事情并非容易。尽管很多学者都对这方面的工作十分热衷，因为作出的成果有可能称为证实或否定半保留复制的重要实验证据。可是却料想不到实验材料相当困难，因为一定要创立一个能量度出个体 DNA 分子微小放射性含量的方法才行。直到 1957 年才获得了方法上的突破。M. Meselson 和 F. W. Stahl 采用稳定的同位素^{15}N作 DNA 标记，而不用^{32}P和^{14}C(放射性同位素)，^{15}N会导致 DNA 分子密度显著增加。这样可以通过密度梯度离心将亲本链和子代链区分开来。他们将 *E.coli* 放在含有^{15}N的培养基上生长，使亲代的 DNA 双链都标记上^{15}N(重链)，若提取 DNA 进行 CsCl 梯度离心应只有一条带位于离心管底部，紫外吸收光谱只出现一个大峰(图 11－2c)。然后再将持续生长的后代放在含^{14}N的培养基中生长，使新合成的链中所含的 N 原子皆为^{14}N(轻链)。按(图 11－2a)若是半保留复制应只出现一种“杂种链”，有^{15}N标记的亲本链和^{14}N标记的子链互补而成；而若是全保留复制应有两种双链，亲代双链皆由^{15}N标记，新合成的链有^{14}N标记；若是按分散型也只会产生^{14}N和^{15}N相同排列的一种双链。实验结果显示离心管中只出现一条带，位于中部，紫外吸收峰也出现一条中等峰。表明半保留复制模型是和实验结果完全吻合，全保留模型可以排除但分散模型却不能排除(图 11－1)；将持续生长的 *E.coli* 再放入含^{14}N的培养基中培养，按半保留复制模型应有两种双螺旋，一种由含$^{14}N/^{15}N$组成的轻链，另一种是由轻链和重链组成的杂种链。这两种链的数量相等；若按全保留复制，预期也会产生两种链，但和前者不同，一种是完全

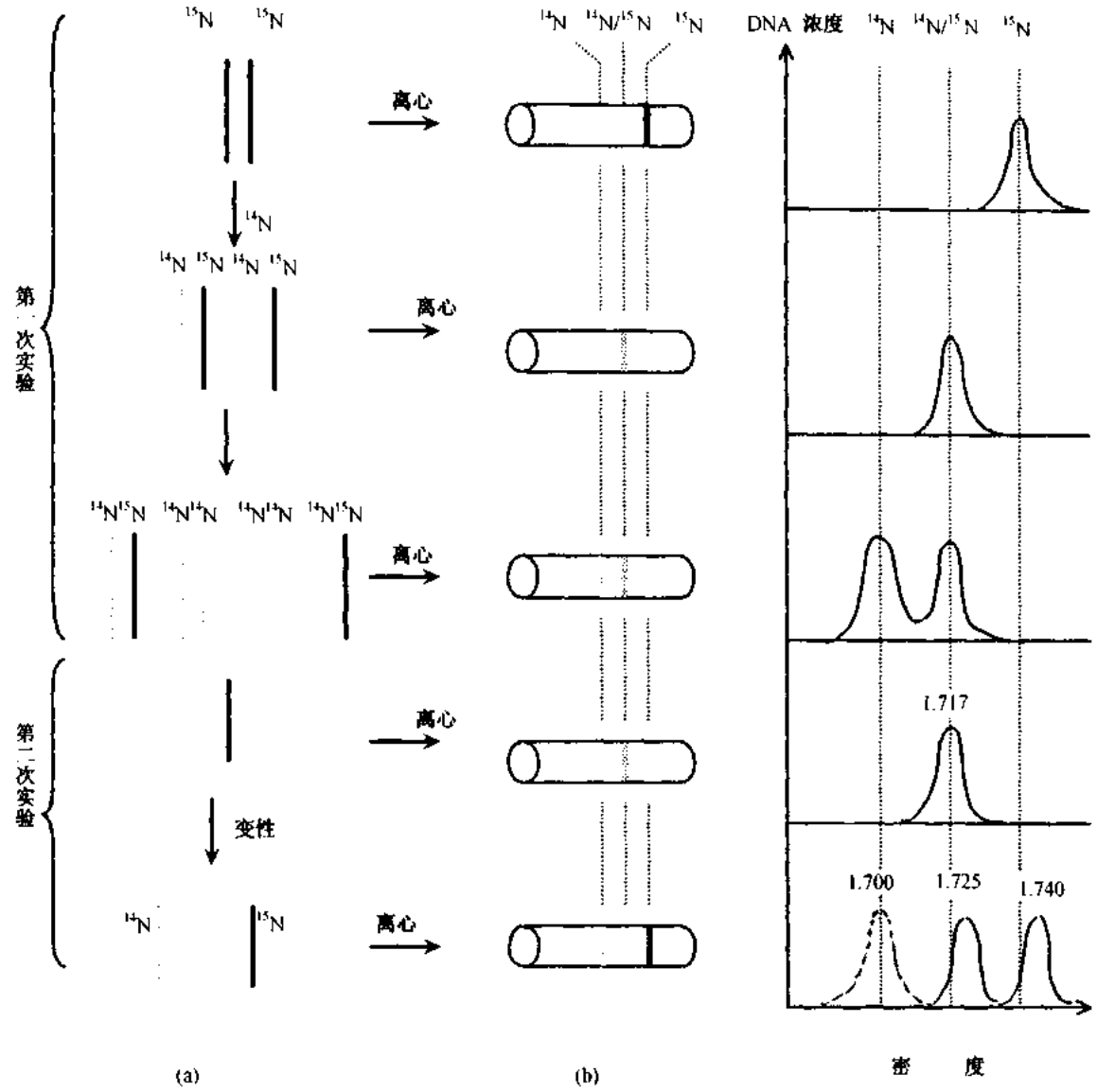


图 11-2 Meselson-Stahl 实验。(a)实验结果的解释;(b)CsCl 梯度离心的结果;离心后 DNA 的吸受率(虚线峰表示对照实验)。

由轻链组成的双链，另一种是完全由重链组成的双链，轻重之比为 3∶1；若按分散模型，第二代子链都是轻重相间排列，仅轻链片段比例要多于重链片段，所以预期仅一条带位于中上部。实验的结果是离心管中持续两条带，比例相等，一条位于上部(轻链)，一条位于中部(杂种链)，紫外吸收也相应持续两个峰，此也完全符合半保留复制，而彻底排除分散模型。虽然按全保留复制也只产生两条带，但比例和位置都不符合，结合第一步实验也可完全排除。从而十分信服地证实了 DNA 半保留复制是完全正确的。

Meselson 等还做了第二项实验，就是将第一代的$^{14}N/^{15}N$ 杂种链经变性将双链分开，然后再将变性前后的双链和单链分别进行 CsCl 密度梯度离心。变性前杂种双链只有一种带，密度为 1.717g/cm^3，变性后两条单链的密度不同，因而呈现了两条带，一条为^{15}N 带，(1.740g/cm^3)，另一条为^{14}N 带(1.725g/cm^3)。而同时进行作对照的样品是从肺炎球菌(*D. pneumoniae*)中提取的 DNA，变性前后都只有一条带，密度为 1.700g/cm^3。

这一结果完全符合半保留复制预期的结果，并彻底否定了“分散模型”和与之类似的末端相连保守模型(end-end conservative)，这两种模型的第一代子链和半保留模型难以区别，但通过变性其两条单链的结构相同，故预期和肺炎球菌一样只能出现相同的一条离心带，但实验结果却是两条带。

二、Taylar 实验

1958 年 Herbert Taylar 用 ^3H 标记蚕豆根尖细胞的 DNA，通过放射自显影证实了真核生物的 DNA 复制也符合半保留模型。

每一条染色单体是由一条 DNA 双链组成，按半保留复制机制，Taylar 预期用 ^3H 标记蚕豆根尖细胞的 DNA 到第二周期后不再用标记，双链复制应如图 11－3(a)所示。经放射自显影后第一周期的中期(M_1)染色体中的两条单体都被标记；到第二周期的中期(M_2)染色体中仅一个单体被标记；到第三周期的中期(M_3)时，染色体有两类：一类是染色单体都未标记，另一类是有一个染色单体被标记，两类染色体比例相等，这一结果(图 11－3(b))和预期的情况完全符合。

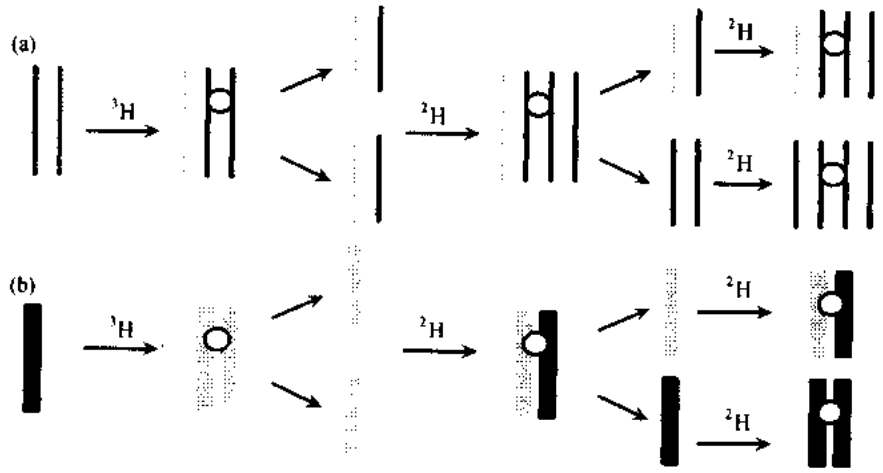


图 11－3　Taylor　蚕豆根尖放射自显影实验。
(a)按半保留复制预期 DNA 在复制过程中的标记情况；(b)实验结果染色体放射自显影的图示。

三、 姐妹染色单体差别染色方法

真核生物的半保留复制也可用现代技术姐妹染色体色差（sister-chromatid differential staining）方法加以证明。在复制过程中用胸苷尿嘧啶的类似物 5－溴脱氧尿嘧啶(5－Bromodeoxy uridine，简称 BUdR)可以掺入到新合成的 DNA 中。无论在何时，新合成链中的 T 被 BUdR 所取代。这样母链中含有 T，而新合成的链中含的是 BUdR。若用荧光染料染色的话，正常链和 T-BUdR 杂种链都可以染色，但含 BUdR 的双链是不能染色的，这样在荧光显微镜下可观察到两条姐妹染色单体的荧光强弱不同。1974 年 Korenberg 和 Freedlender 改进了这一技术，在掺入 BUdR 后经紫外光的照射，可损伤含有 BUdR 的双链，这样再用 Giemsa 染色，只有含 T 双链或 T-BUdR 杂种链可被染色，而合 BUdR 双链则因 DNA 破坏而染不上色。这样在培养的第三周期中期染色体的两个姐妹染色单体不同，一个染上深紫色，另一个不染色，故称之为**斑色染色体**(harlequin chromo-

some)(图 11－4,11－5)。和前面的道理一样,这种差别染色也能很好地证实真核 DNA 的半保留复制。现在这种方法又被应用于三致物质(致癌变、致畸变、致突变)的检测,方法是通过第三周期中期染色体的姐妹染色单体交换(sister chromatin exchage)频率来决定。

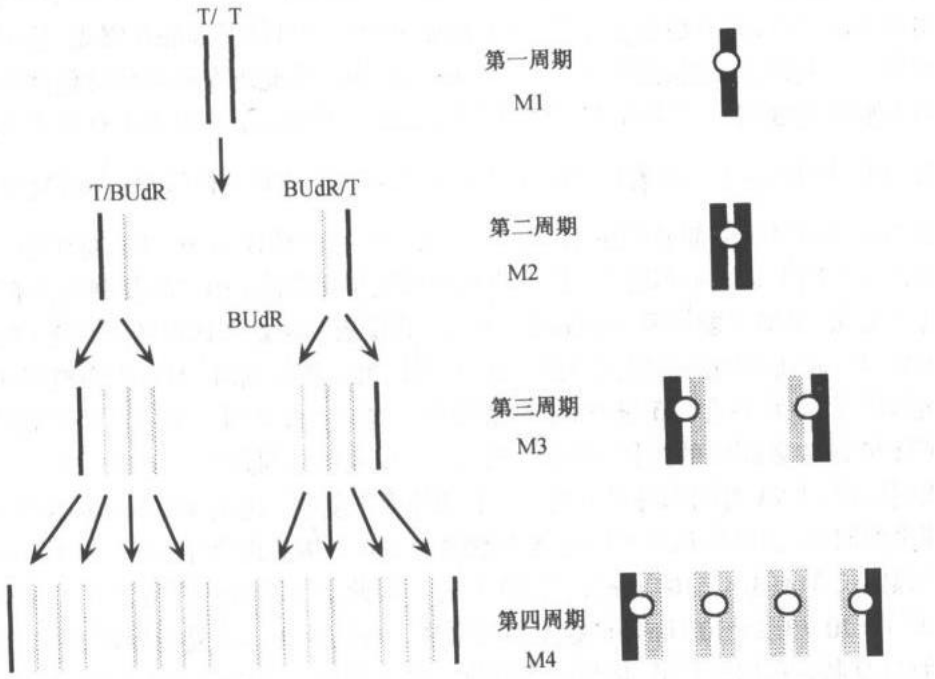


图 11－4 姐妹染色单体差别染色的原理和结果。

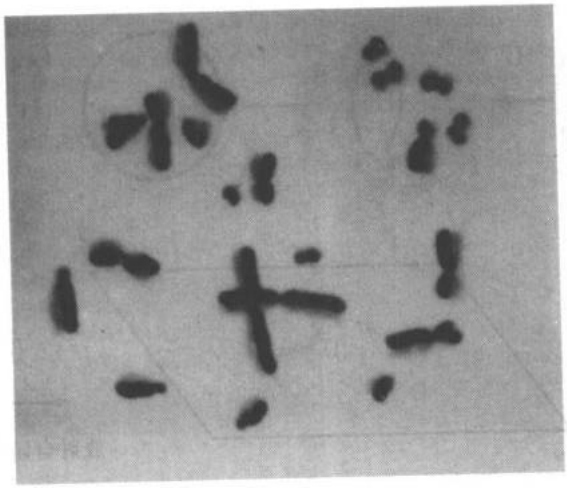


图 11－5 中国仓鼠的斑色染色体。(引自 Russell,P.J.:Genetics,3rd ed.,1992,Fig.11－20)

四、Cairns 复制模型

Meselson 和 Stahl 的实验虽然有力地证明了半保留模型,但还有一些问题需要解决:在原核基因组中复制是沿着环状 DNA 先后逐步完成的,还是在很短的时间里同时完成的?显然 DNA 合成是一个逐步的过程,那么为什么实验中未出现某一片段上只有一半复制了,还有一半未复制呢?即存在轻的和重的分子联合存在的片段。为了解决这

一问题，1963 年 Cairns 设计了另一实验。他将细菌 *E.coli* 放在含有标记的 T 的培养基中培养，新合成的 DNA 链中含有放射性，按半保留复制模型，培养一代后产生的环状子染色体中应有一条单链是被标记的。那么继续在含培养基中培养 1 代时，第二轮复制已经开始，每个环状染色体含 ^3H 的部分如图 11－6 所示，此时将溶菌酶缓慢裂解细胞壁，以提取完整的环状 DNA 放到透析膜上，再固定在载片上，罩上照相底片后置于暗处放 2 个月，进行放射自显影，由于释放出的粒子可使照片底片曝光，显示出标记的部分。在电镜下观察、摄影，获得了 θ(theta)结构。因此现在对环状染色体的半保留复制也称 **θ 型复制**。实验结果首先证明了 Jacob 和 Wollman(1957 年) 有关 *E.coli* 基因组为圆形排列的推断。放射自显影照片还表示 *E.coli* 环状染色体在 $1\frac{1}{2}$ 时复制的情况。第二轮尚未复制的部分(ABC)已标记了一条链，叫做**热**(hot)**链**，未被标记的是**冷链**，放射自显影图中是显示不出的。这表明复制却是半保留的。第二轮复制时只有部分环进行了复制，形成了**复制叉**(replication fork)，其模板一条是未标记的亲代链(AEC)，另一条是第一轮新复制的被标记的链(ADC)，所以第二轮复制时，ADC 两条链都被标记过，显影较深，AEC 只有一条链含 ^3H，显影较浅。这样不仅证明了半保留的正确性，还显示了环状 DNA 复制时的拓扑学形态(图 11－6)，表明复制是有一定序列的，在一轮复制未完成前，DNA 有的部分已复制，有的部分尚未复制，这是 Meselson-Stahl 实验所没有解决的问题。用这张照片可以直接度量出 *E.coli* 基因组的大小(全长为 1100μm)，而每个螺旋长 34Å($3.4\times10^{-3}\mu m$)，含 10 个 bp，因此 *E.coli* 的基因组含有 $1100\div(3.4\times10^{-3})\times10=3.2\times10^6$(bp)，根据实验结果进一步推论 *E.coli* 复制周期有一个固定的起始点；复制是单向进行的。以后的实验进一步证明第一个推论是正确的，而第二推论却未能言中。

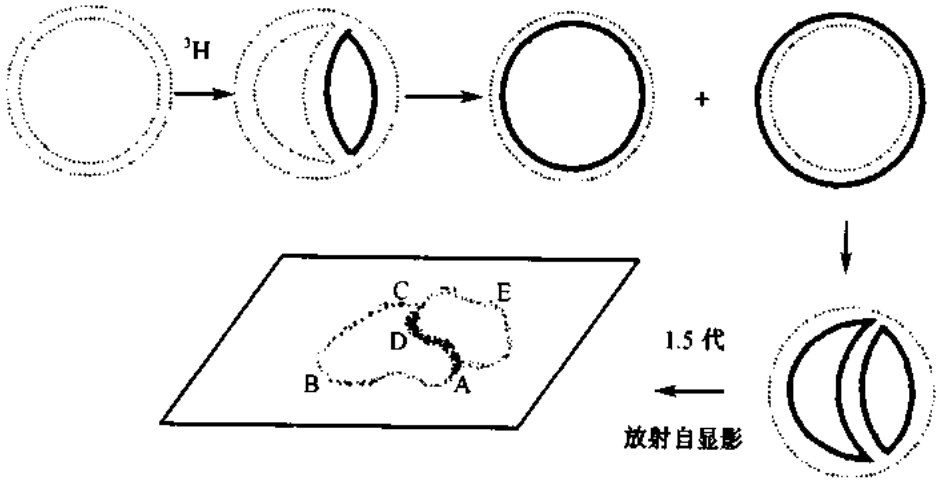


图 11－6　Cairns 实验原理及过程。

第二节　复制的起点、方向和终点

Cairns 推理 *E.coli* 的复制有起始点，且单向复制。他的推理是否正确呢？实验上从 Cairns 模型中尚不能信服地解决这些问题。从 θ 模型可以推测有复制起始点，但进一步要问这个起始点是否固定？在染色体的什么位置上？作为复制起始点有无特殊的序列和

结构？至于复制的方向问题，从复制叉来看应有两种可能？单向和双向，具体确定必须要有进一步的实验证据。人们采用了各种巧妙的方法将这些问题都一一化解了。

一、研究方法

1. 用同位素标记电镜观察及变性法

用同位素标记法有两种思路，一是假设若有一个起始点，且双向复制那么起始点和复制的终点相对排列；若只有一个起始点，单向复制的话，起始点和终点重合（图 11－7）。因此在第一周期末开始标记至第二周期始，实验结果表明第二周期细胞的环状染色体有半数上面有两个相对排列的标记位点（图 11－8），这就是**复制起点**和**终点**，表明 *E. coli* 有一个复制起点，且为双向复制。这是由 Rodriguez 提出的，为了保证同步复制，他选择了温度突变型（42℃只能延伸不能起始），先将细菌在 42℃下培养，直到复制停止，再放到 25℃下培养，这样同步进入新周期。另一个思路就是仅在复制开始时进行轻标记，然后继续再进行重标记，如果两个 Y 型复制叉都被标记，说明复制是单起点，双向复制，若仅一个复制叉被标记说明单向复制，轻重标记可以显示复制的先后序列（图 11－9）。

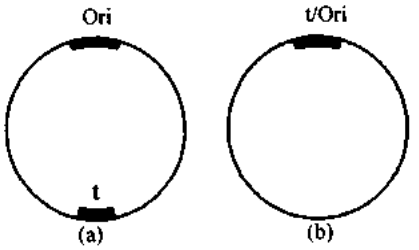


图 11－7 (a)若有固定起始点，双向负制，则复制终点应在起点的对面。(b)若有固定起始点，单向复制，则复制终点应和起点重叠。

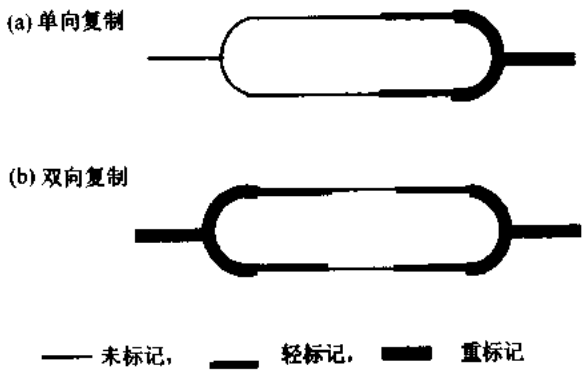


图 11－8 用同位素标记复制叉来确定是单向复制还是双向复制。(仿 B.Lewin: Genes, Ⅵ, 1997, Fig14－5)

(1) 用噬菌体插入标记法

这种方法是由 Bird 等于 1972 年建立的。**Mu 噬菌体**是一种可引起突变的噬菌体（见第 16,23 章），它可以随机插入到大肠杆菌基因组中的多个位点中，当它插入到某一个基因内，这个基因就失去了功能，使细胞突变成了缺陷型，因此我们通过选择培养基可以将突变型选出来。Mu 噬菌体的结构和序列也十分清楚，用其基因组中的特殊序列作为探针，进行分子杂交能确定其插入的相对位置。前

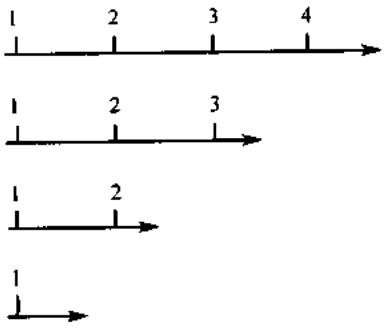


图 11－9 不同步培养时各标记的拷贝数不同，复制起点标记的拷贝数应最多。

面已介绍λ噬菌体只在 *att* 位点整合到的 *E.coli* 染色体中，整合时不会导致任何基因的突变，而且每个 *E.coli* 染色体中仅有 1 个拷贝，因此，λ 的拷贝数可以用来表示复制的周期数。有了这两项前提就可以采用**同步培养**和**不同步培养**来测定复制的起点和方向。所谓同步培养法就是将各种噬菌体的插入突变型同时培养，若有固定复制起始点，那么这一点首先获得 2 个拷贝。不同步培养是将不同突变体群体进行培养，在不同时间里抽样检查，观察那个突变型复制的拷贝最多。若有固定点则毫无疑问，在不同的时间里它都能得到复制，而远离起始点的位点在短时间就得不到复制，所以复制的拷贝数可以表示复制的起点和序列。这种方法不仅解决了有无复制起点及是否单双相复制的问题，还确定了具体的复制起点和终点。

(2) 变性定性法(denaturation mapping)

是由 Ross Inman 和 Maria Schnös 建立的。方法是将 DNA 分子加热到刚能出现解链的温度时，A－T 丰富区便解链形成“小泡”，若加入变性剂甲醛，并使 DNA 冷却，小泡依然存在。这样不同位置上的小泡就成了可以用电镜观察到的标记，作为固定的参考点。复制时就可以根据复制叉和小泡的相对位置来确定复制起始点和复制方向(图 11－10)。

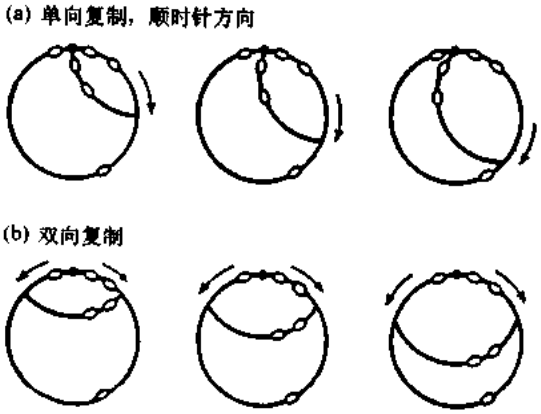


图 11－10　复制叉的移动和小泡。(引自 D.Freifelder: Molecular Biology, 1983, Fig. 8－44)

(3)双向电泳法

是现代测定某个基因与复制起始点的距离的方法。现假设有基因 1,2,3 在复制起点一侧按次排列，在 1－3 外侧各有一个限制性酶切位点。现将此基因组在体外进行不同步复制，那么就会产生不同长度的复制叉 (图 11－11(a))，然后用限制酶 R 降解基因组，再通过柱层析从各种不同长度的 DNA 中富集含基因 1、2、3 的片段，再进行双向电泳。第一相为中性琼脂糖(1%)电泳，Y 型复制叉的长短不同，泳动的速度不同，Y 复制叉越短，移动越快，这样不同长度的 Y 复制叉走成连续的一条带(图 11－11(b))。将这条胶整齐割下，放在电泳板上作为负极再灌第二相碱性琼脂糖(1.5%)胶(40mmol/L NaOH, 2mmol/L EDTA)，继续电泳。由于在碱性条件下双链变性，这样 Y 叉中新合成的不同长度的单链就会游离出来，这样原来的模板链各起点都是等长的，且长于任何新合成片段，所以跑得最慢靠近负极，原连续排列的长度不同的新合成链的长度不同，在胶上以弧形排列。把此胶移到尼龙膜上，分别用 1、2、3 三个基因的探针来进行杂交，由于基因 1 最靠近起点，新合成链无论长短，大部分有此序列的存在，故杂交的阳性弧形带最长。3 基因离复制原点最远，只有 Y 复制叉达到相当长时，此基因才可能得到复

制，所以其杂交的阳性弧形带最短(图 11－11(c))。同理 2 基因的杂交阳性弧形带长度为中等。若是未知序列的基因，我们只要通过此实验以杂交带的长短来判断它们离复制起点的远近而进行排列。

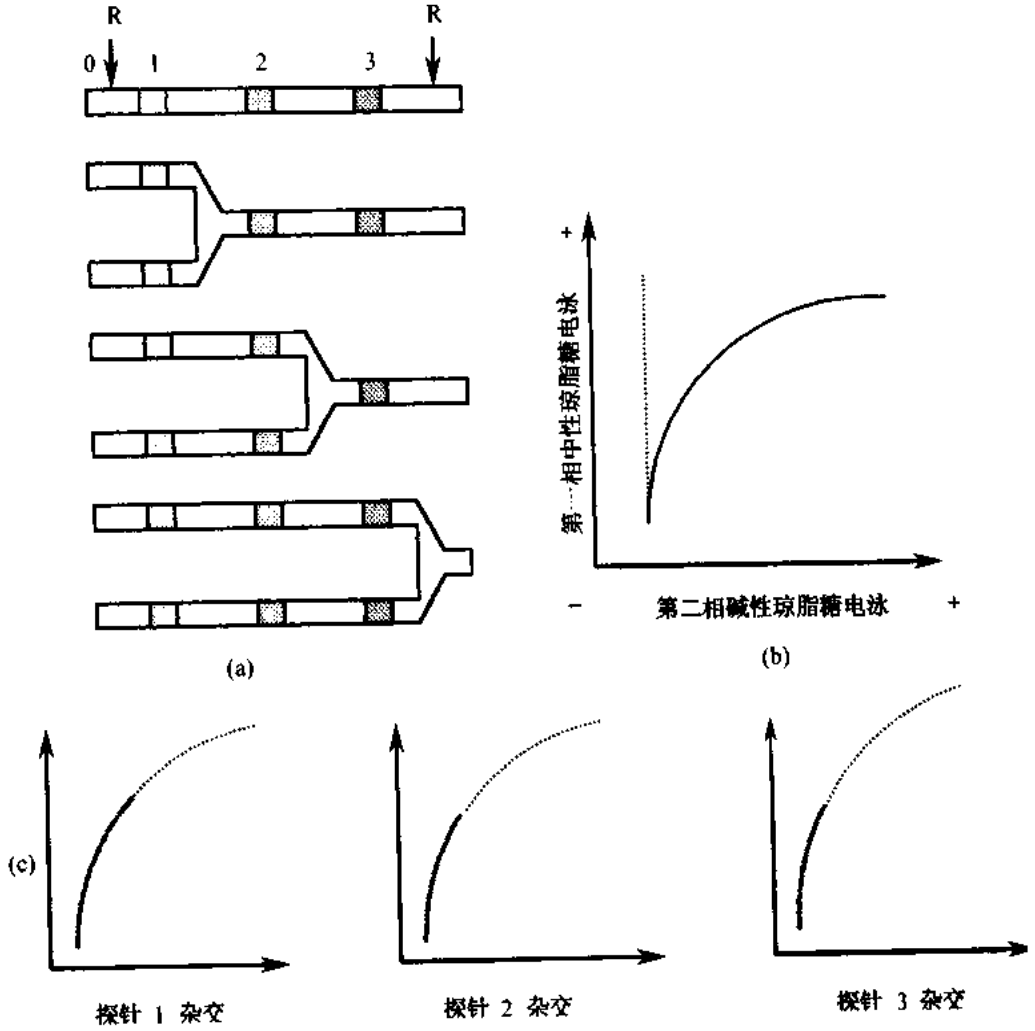


图 11－11　双向电泳定位法。

以上的各种实验各有利弊，但思路都十分巧妙，电镜法和变性法，方法简便，能确定有几个复制起点，是单向复制还是双相复制，但不能具体确定起点和终点在何处(在什么基因位点)。插入突变法，工作量大，过程繁杂但不仅能确定起点的有无，是否是单、双相复制，还能确定具体的位点。变性法能确定紧密连锁的几个基因和复制点的距离和排序。但也不能确定起始点的位置。

二、原核的复制起点和方向

现在通过各种实验表明，大肠杆菌有一个复制起点位于 74.5′的 *ilv*(isoleucine－valine)操纵子位点，终点在 25′的 *Trp* 操纵子位点附近，属双向对称复制。T7 是线性染色体，也只有一个复制起点，在近一端的 17%处开始，然后向两端延伸。真核的复制和原核不同，有多个复制起点(i)，双向等速复制。

在原核生物也并不都是定点、双向对称的模式，也发现一些例外的情况。如枯草杆

菌的复制虽有固定的起始点，双向复制，但是是不对称的。一侧复制基因组的 $^4/_5$，另一侧复制基因组的 $^1/_5$，即终点不在 ori 点的对面(图 11 - 12(a))。质粒 R6K 早期为单向复制，复制了约 $^1/_5$ 基因组时，进行双向的晚期复制(图 11 - 12(b))。质粒 Col E1 有固定起始点，但却为单向复制(图 11 - 12(c))。

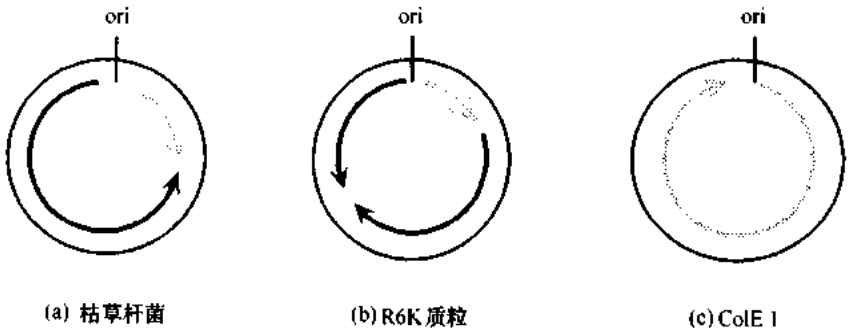


图 11 - 12　原核生物中特殊的复制类型。

真核生物的线粒体按共生学说其来源于原核，它的染色体也是环状的，但复制十分特殊(图 11-13)。哺乳动物的 **mt DNA** 有两个复制起点，一个在重链(H)上，另一个在轻链(L)上，两者相距较远。H 链上的复制点先开始起始复制，按 5′→3′方向合成新的 **L 链**，但不稳定，短链不断降解、合成，维持一个开放区。5′仅有一个，3′端有 3～4 种不同的位点。这部分的亲代 L 链被新 L 链所取代，形成一个取代环(displaced loop)，也叫 **D 环**，长 500～600bp，随着 L 链的合成，D 环不断扩展。当取代环达到 mt DNA $^2/_3$ 处时，L 链上的 ori 露出，新 H 链开始合成。新的 L 链完全合成后，释放出新代的 L 环，此时亲代 L 环部分进行了复制。经继续合成最终也完成复制产生带一个缺口的双环，两个新的 L 链和 H 链的缺口被封闭起来，完成全部复制。有的生物如爪蟾的线粒体只有一个长的 D 环，而另一些生物有几个短的 D 环，如四膜虫线性 mt DNA 就有多达 6 个 D 环。在高等植物的叶绿体中有 2 个 D 环。

第三节　DNA 复制突变型的筛选

为了进一步研究复制过程有哪些酶和蛋白质参加，人们就采用筛选各种突变型的方法。DNA 复制的各种突变型都导致不能合成 DNA，最终致死，但可分为两类，一类是可以补偿的，这种突变型是由于某些基因的突变不能合成某种核苷酸所致。只要我们给它加入所缺乏的原料 - 核苷酸便可顺利合成 DNA。另一类是不可补偿的，这类突变型可能是不能合成某种复制所需的蛋白质因子或酶，因此难以补偿。若能分离研究这些突变型，就能使我们了解复制过程中需要哪些蛋白和酶，以及它们的结构和功能。因为它们是致死突变，所以只有选择条件致死的品系才能将它们保存下来加以研究。一般有以下几种筛选方法。

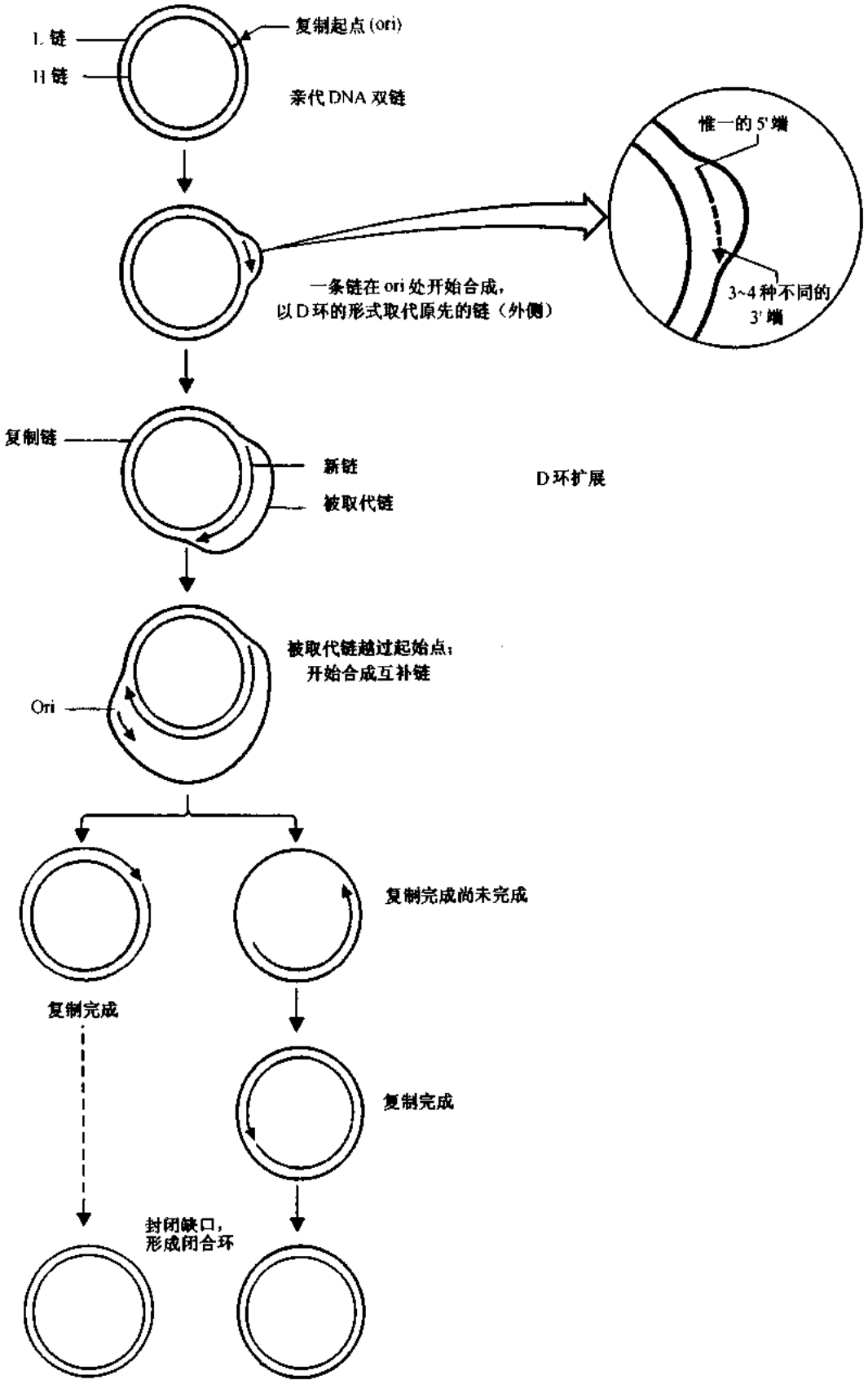

L 链
H 链
复制起点 (ori)
亲代 DNA 双链
惟一的 5' 端
3~4 种不同的 3' 端
一条链在 ori 处开始合成，
以 D 环的形式取代原先的链（外侧）
复制链
新链
被取代链
D 环扩展
被取代链越过起始点；
开始合成互补链
Ori
复制完成尚未完成
复制完成
复制完成
封闭缺口，
形成闭合环

图 11－13　D 环复制。

1. 根据温度敏感性筛选

将诱变剂处理细菌，引起突变，然后接种平皿在低温（30℃）条件培养，然后再影印到另一培养皿中，在 42℃条件下培养，不能长出的菌落是温度敏感性突变。用这个方法筛选出了 *dnaA, B, C, D, E, F, G7* 个与复制有关的基因。

2. 同位素标记法筛选

温度敏感突变型有的和复制有关，有的和复制无关，为了加以区分，可采用同位素标记法。用诱变剂处理细菌，先在低温培养，然后影印放在高温下培养，培养基中含有用^{14}C 标记的 T，通过放射自显影未能掺入^{14}C 的菌落为复制突变型，用此法筛选出 *dna E, B, G* 突变型。

3. 5－溴尿嘧啶掺入法

和上面同样的原理，但可不用同位素，而用 5－BUdR。可将被诱变的细菌放在高温条件下培养，培养基中加入了 T 的替代物 5－BUdR，仍能复制的细菌就会从培养基中吸收 5－BUdR，在 313nm UV 的照射下，因掺入 5－BUdR 的 DNA 对 UV 敏感而使细菌被杀死，而复制突变型得以存活。

4. 质粒温度敏感性的筛选

此是 Jacob 等（1963）建立的方法，质粒的复制和宿主的复制机制有关，因此筛选出质粒的复制突变型也能用于研究复制所需的蛋白和酶。人们用 *E.coli* F-*lac*（lac^-/lac^+）品系，即 *E.coli* 本身为 lac^-，它带有一个含 lac^+的 F′ 因子。将此细菌诱变，在 EMB（伊红美蓝培养基上，当基因正常时可引起发酵，在红色培养基上，阳性菌落呈深紫色）培养基上（30℃）培养，菌落为深紫色。然后将它们影印到 EMB 培养基上，在 42℃培养，若菌落不是深紫色表明基因 *lac* 发生了突变，失去了功能。那么进一步要问这种突变是由于 F′因子不能复制所致，还是基因 *lac* 本身功能发生改变（即高温时不表达、低温时才表达）引起的呢？我们只要将在 42℃含非深紫色平皿，放回 32℃培养，若仍然是非深紫色表明是复制引起的。因当 F′因子不能复制时宿主照样繁殖，这样大量的后代中已不再带有 F′lac^+，变成 lac^-品系，所以温度降下来也无济于事。但若属于温度敏感性功能的变化，那么质粒仍然能复制，每个细胞中都还含有 F′ lac^+，只是在低温中才表达而已，用这种方法分离出了 *dna N* 的突变型。

5. 根据抗药性筛选

有的抗生素如新生霉素（novobiocin）可和与复制有关的蛋白（Top Ⅱ，β 亚基）结合而抑制复制。*E.coli* 温度敏感型中有的是编码拓扑异构酶 β 亚基的基因发生了条件突变，在提高培养温度时，这种产物的结构发生改变，不再和新生霉素结合，结果产生了抗药性。通过这种特异性的筛选就可鉴别出拓扑异构酶的 β 亚基。

6. 根据与突变有关的生物学功能筛选

前面已介绍了细菌的接合，接合的过程必须伴随着复制，在丧失接合能力的突变型中，有的是因不能正常复制而产生的，如从突变型中分离出了编码腺苷激酶的基因 *dna W*，此基因和 DNA 前体合成有关。

7. 快停突变与慢停突变

复制的过程总的说来包括起始、延伸和终止三个阶段，一轮一轮的进行，每一阶段的蛋白质和酶不完全一样，有的是某一阶段所特有的，我们如何把这种蛋白质或酶的基因筛选出来呢?**快停突变**(fast-stop mutants)和**慢停突变**(slow-stop mutants)就是一种较好的筛选方法。所谓的快停突变，就是当把培养温度提高以后(42℃～45℃)复制迅速停止(图 11－14)，说明这种突变所涉及的产物和链的延伸有关而和起始无关。慢停突变是指将培养温度提高以后，DNA 合成并不立即停下来，还会继续合成，要经过一段时间，合成才会停止。为什么会产生这种情况呢? 那是因为复制是一轮一轮地进行，如果第一轮已经起动了，这时将温度提高若引起了与起始复制有关蛋白和酶的突变，并不会影响 DNA 的延伸，所以合成继续进行，直到第二轮合成起始时才会因某种蛋白或酶的功能改变而不能进行，使 DNA 合成停下。这样就可以筛选出与合成起始有关蛋白或酶的基因。

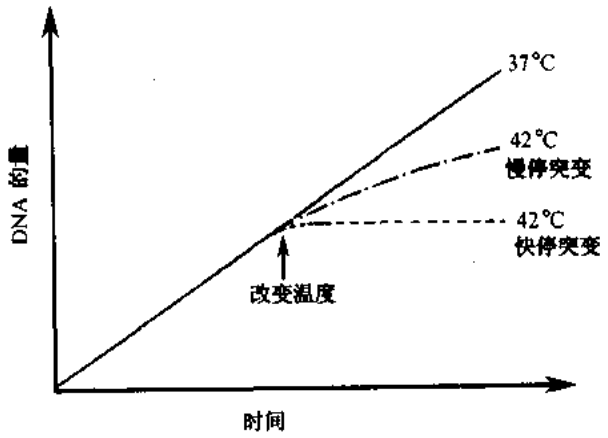


图 11－14　慢停突变和快停突变示意图。当提高培养温度时快停点变的 DNA 合成立即停止，而慢停突变则缓慢停止。

第四节　原核生物复制的酶系统

一、DNA 合成酶

在 20 世纪 50 年代的中期，A．Kornberg 和他的同事们就想到 DNA 的复制必然是

一种酶的催化作用，于是决心分离出这种酶并研究其结构和作用机制。为了达到这个目的，他们分离 *E. coli* 的蛋白，然后加到体外合成系统中，即同位素标记的 dNTP、Mg^{2+} 及模板 DNA，经过大量的工作，于 1956 年终于发现了 **DNA 聚合酶** Ⅰ（DNA polymerase Ⅰ，DNA pol Ⅰ）原来称为 Kornberg 酶。以后又相续发现了 DNA pol Ⅱ和 DNA pol Ⅲ。开始人们以为 DNA pol Ⅰ是细菌中 DNA 复制主要的酶类，后来发现 DNA pol Ⅰ的突变株照样可以复制，才清楚它并不是主角。现在已经知道在 DNA 复制中起主导作用的是 DNA pol Ⅲ，至于 pol Ⅱ的功能现在还不十分清楚。DNA 聚合酶的共同特点是：①需要提供合成模板；②不能起始新的 DNA 链，必须要有引物提供 3′－OH；③合成的方向都是 5′→3′；④除聚合 DNA 外还有其他功能。

所有原核和真核的 DNA 聚合酶都具有相同的合成活性，都可以在 3′－OH 上加核苷酸使链延伸（图 11－15），其速率为 1000 nt/min。加什么核苷酸是根据和模板链上的碱基互补的原则而定的。

E. coli 的 DNA pol Ⅰ涉及 DNA 损伤修复，在半保留复制中起辅助的作用。DNA pol Ⅱ在修复损伤中也是有重要的作用。DNA pol Ⅲ是一种多亚基的蛋白。在 DNA 新链的从头合成（*de novo*）中起复制酶的作用。三种酶的结构与活性如表 11－1 所示。

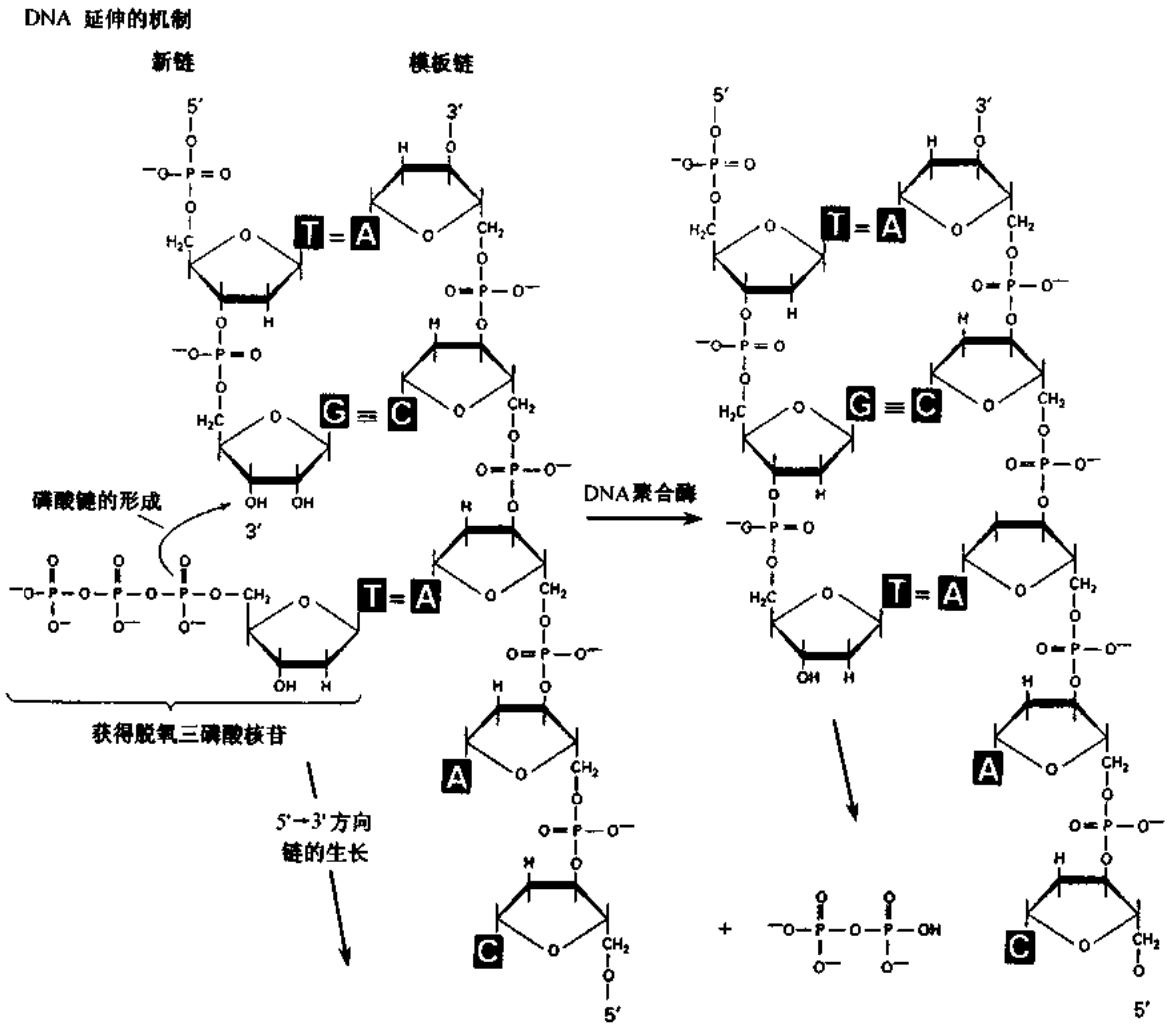


图 11－15　DNA 聚合酶的合成机制。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 11－13）

表 11－1 *E.coli* 中的三种 DNA 多聚酶

DNA pol Ⅰ		DNA pol Ⅱ	DNA pol Ⅲ
结构：			
分子量	109kDa	90 kDa	900kDa
构成	单体	单体	异多聚体
分子数/细胞	400	?	10～20
酶活性：			
5′→3′聚合酶	+	+	+
3′→5′外切酶	+	+	+
5′→3′外切酶	+	－	－（可切单链）
突变性：			
突变位点	*pol* A	*pol* B	*pol C*(*dna E*)，*dnaN*，*dnaZX*，*dnaQ*，*dnaT*
突变表型	修复有缺陷	能修复	阻止复制
条件致死	仅5′→3′外切活性受影响时，活率下降	不受影响	条件致死

摘自 Gene Ⅴ，1994，B.Lewin.Table 19.1

复制的忠实性问题会影响到翻译的精确性。这种忠实性主要是依赖于碱基的特异性配对。但估计每个碱基对将有 10^{-3} 的机率会发生错配，但实际的错配率只有 10^{-8}～10^{-10}，即每 1000 细菌复制周期中，每个基因组只产生一次错误。DNA 多聚酶能增加互补碱基特异性，选择的特异性主要是两方面的作用。①它能够检查进入的碱基是否和模板互补，这时通过一个匹配的化学特点加以识别，这是合成前的一种预防措施，称为**合成前**(presynthetic)**错误控制**。②在新的碱基加到链上以后，来检查碱基是否配对。若发生错配，将会把刚加上去的错误碱基切除掉。这是**校正控制**(proofreading)。

细菌的三种 DNA pol 都具有 3′→5′外切酶活性，逆 DNA 合成方向进行加工，提供校对功能(图 11－16)。在链的延伸阶段中，一个核苷酸进入生长链的末端，形成键，这酶就向前移一个碱基对，准备下一个碱基对的进入。若一个错误产生了，这个酶就向回退，切除最后加进来的这个碱基，产生一个位点，为正确的碱基取代。

不同 DNA 聚合酶聚合和校正之间的关系是不同的。有时，这些活性是由相同的蛋白亚基产生的，但有时它们又还有不同的亚基（表 11－2)。一个错误碱基的排除实际上是十分复杂的。因为这种切除反应是由不同位点催化的。

1. DNA 聚合酶 I 的结构

DNA pol Ⅰ是一个多功能的酶，有 pol A 编码的单个肽链的分子，分子量为 103kDa，枯草杆菌蛋白可将其切成两个片段，小片段位于 N 端，分子量为 35kDa，具有 5′→3′外切酶活性。大片段位于 C 端，分子量为 68kDa，称为 **Klenow 片段**，其 C 端 $^2/_3$ 区域是含正电荷的直径为 2nm 的裂缝，具有聚合活性，DNA 双链就结合在裂缝中，一旦结合，随即裂缝就关闭起来，此时 DNA 只能在其中滑动。余下 $^1/_3$ N 端区域具有校正的外切酶活性，这两区域间隔 3nm(图 11－17)。

DNA 聚合酶有 6 个结合位点：①模板结合位点；②引物结合位点；③引物 3′－OH 结合位点；④底物 dNTP 结合位点；⑤5′→3′外切酶结合位点；⑥3′→5′校正位点。DNA

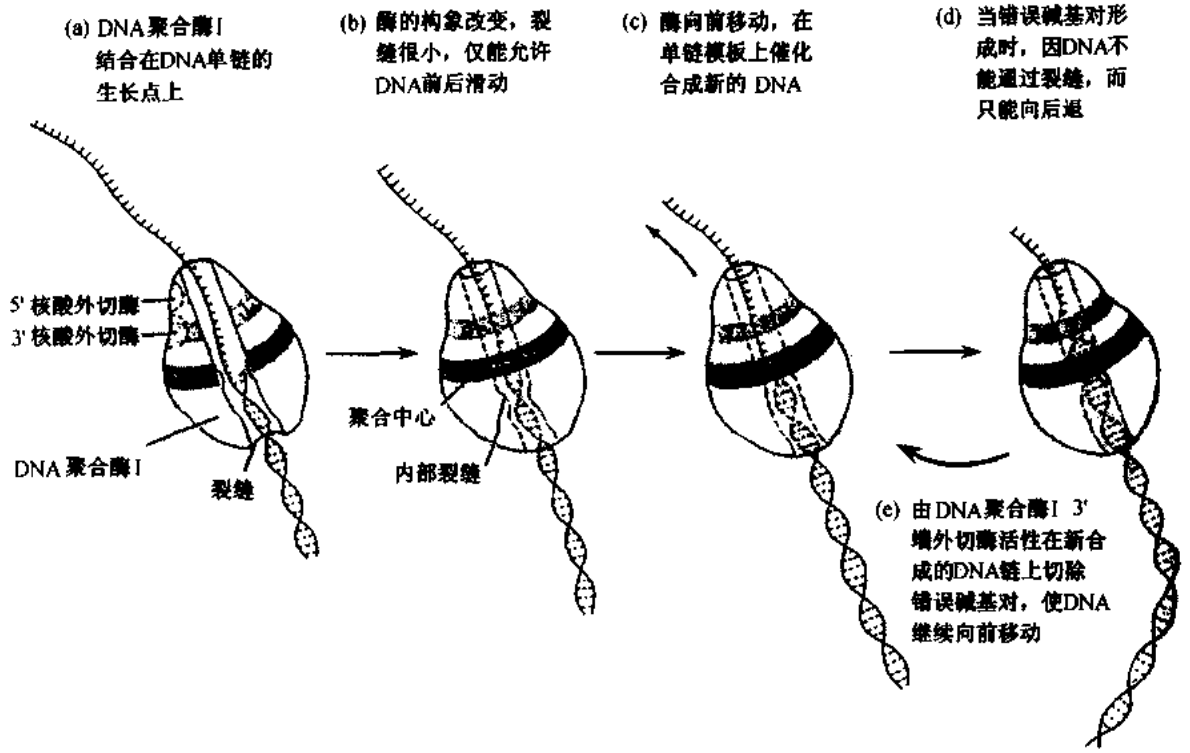


图 11－16　当错误碱基时人 DNA 后 DNA 聚合酶 I 的校对作用。(转引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 11－13)

表 11－2　DNA 多聚酶的校正功能

酶	合成功能区	校正功能区	错误率	
			校正前	校正后
E. coli DNA pol Ⅰ	aa 200－600	N 端	10^{-5}	5×10^{-7}
E. coli DNA pol Ⅲ	α 亚基	ε 亚基	7×10^{-6}	5×10^{-9}
T4 DNA pol	C 端	N－端	5×10^{-5}	10^{-7}
T7 DNA pol	?	118－148	10^{-5}	10^{-6}
反转录酶		没有	10^{-5}	

摘自 Gene V, 1994, B. Lewin. Table 19.2

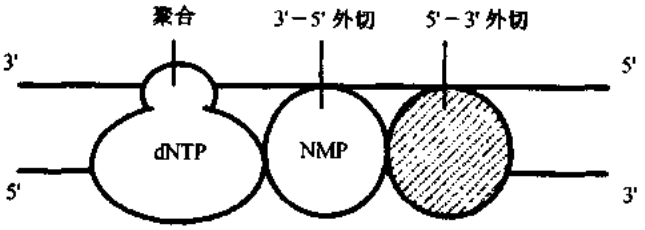


图 11－17　DNA 聚合酶Ⅰ的结构和功能。

pol Ⅰ在不同的位点执行着不同的功能。

2. DNA 聚合酶 I 的功能

(1)5′→3′聚合功能

DNA pol Ⅰ的聚合活性和其他的 DNA 聚合酶的特点相同，但主要用于 DNA 的修复和后滞链中 RNA 引物的置换。它使用模板的范围较宽，有三种：①有引物的 ssDNA（无论线状或环状）；②有缺口（无论大小）的 dsDNA；③有切口（nick）的双链 DNA。而 DNA pol Ⅲ只能利用 < 100nt 缺口的 dsDNA 作模板。

(2) 3′→5′外切活性

DNA pol Ⅰ C 端有一纵沟，沟底由六个反向平行的 β 折叠构成，宽约 2.0～2.4nm，沟的二侧面由 α 螺旋构成，深约 2.5～3.0nm，可以容纳一条 DNA。聚合活性区域在 C 端，3′→5′外切活性区域在中间部位。在新合成的 DNA 链中，错误的碱基即使是蒙混过关进入 DNA 中，但插入的速率很低，同时也不能形成正确的氢键，使双螺旋发生形变，直径加大，DNA 不再能在沟中向前滑动，只有倒退，错误碱基正在位于 3′→5′外切活性功能域，它激发了此区 3′→5′外切活性，将其切除，转换成游离的 dNMP，“障碍”排除后，DNA 继续向前移动，合成新链（图 11－18）。

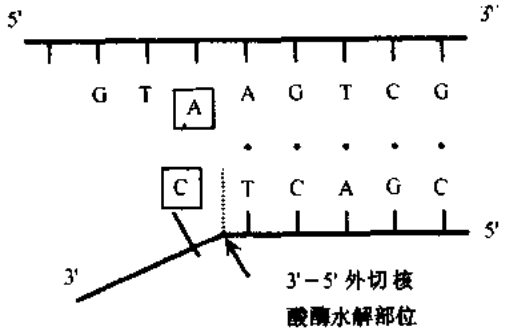


图 11－18 DNA 聚合酶 I 的 3′－5′外切酶活性。（仿 D. Freifelder: Molecular Biology, 1983, Fig. 8－16）

(3) 5′→3′外切活性

此是由 35kDa 小片段执行的 5′→3′外切功能，切除小片段 DNA，一次切除 10 个碱基。此是和合成、校正活性相一致的。这种活性为 DNA pol Ⅰ提供了在有缺刻的 DNA 上开始体外复制的能力，在磷酸二酯键断裂处，此酶可以延伸 3′－OH，从而可以进行 **①切口平移**（nick translation）（图 11－19a）；**②链的置换**（图 11－19b）；**③模板转换**（template-switching）（图 11－19c）。由于 DNApol Ⅰ既具有 5′→3′的聚合活性，又有 5′→3′的外切活性，所以它可以一边从缺口的 5′－P 开始行使外切功能降解 DNA（图 11－20），同时从缺口的 3′－OH 开始合成新链，填补缺口，使“缺口”像是从 5′→3′不断在移动，反应停止后留下的缺刻由连接酶来封闭。这一功能常出现在 DNA 损伤修复、21 后滞链中引物的切除。在体外也常用来作为 DNA 标记的一种途径。当控制酶的反应条件使其 5′→3′外切活性消失或者用 Klenow 酶在具有缺口的双链 DNA 模板上合成时，因它们已不具外切活性，因此产生了链的取代和模板转移。链的置换被认为是重组机制中的重要一环（见第 23 章）。在 *E. coli* 中只有 DNA pol Ⅰ具有独立的链置换能力。在其

他蛋白质的帮助下，前导链可进行链的置换(置换了互补链)，但后滞链中无此现象。模板转换未发现生物学意义。

(4)内切酶活性

当 DNA 存在不配对碱基时，DNA pol Ⅰ具有内切酶活性，在不配对碱基的 5′－P 端的后面切开磷酸二酯键，主要存在于切除修复系统中(见第 21 章)，在正常复制中无此活性。

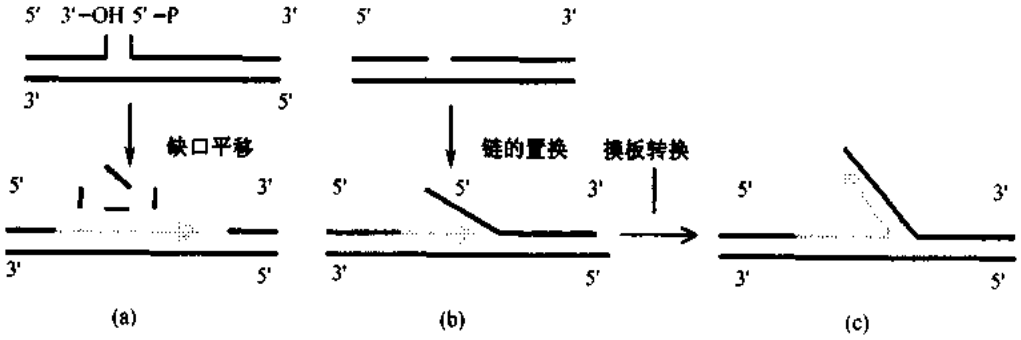


图 11－19　DNA 聚合酶Ⅰ5′－3′外切活性的功能。

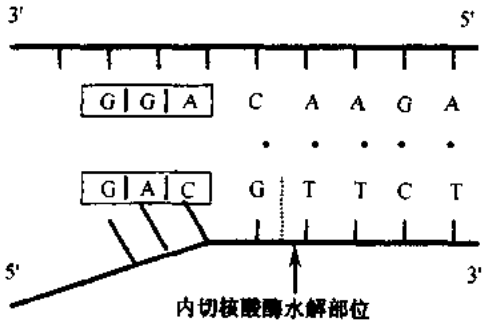


图 11－20　DNA 聚合酶Ⅰ的内切酸活性。(仿 D. Freifelder: Molecular Biology, 1983, Fig. 8－20)

3. DNA 聚合酶Ⅲ的结构

DNA 聚合酶Ⅲ相当复杂，其结构和各亚基的功能如图 11－21,22 所示。

DNA pol Ⅲ具有 5′→3′聚合功能，但模板的要求很高，仅有缺口 < 100b 的 dsDNA 才可做模板。3′→5′的外切活性和 DNA pol Ⅰ相同，有校对的功能。但 5′→3′的外切活性不同，仅可以单链为底物，因此不能进行缺口平移。DNA pol Ⅲ是合成 DNA 的“主角”，一般合成长片段，如前导链的合成和冈崎片段的合成。

现在对 DNA pol Ⅲ有了进一步的了解。它的结构和半不连续复制的机制是相吻合的。其“核心酶”含有 α，ε 和 θ 三个亚基。α 亚基具有合成 DNA 的能力，ε 亚基具有 3′→5′外切酶的功能，起到校正的作用。而 θ 亚基可能在组装中发挥功能。但核心酶的进行性是低的，通常使合成 11 个碱基左右就从模板上解离下来。

α 亚基加到核心酶上有助于使其二聚化，进一步产生 pol Ⅲ*，γ 复合体加入到 pol Ⅲ*，进一步形成 pol Ⅲ′ 复合体。γ 复合体是由 γ，δ，δ′，χ 和 ψ 亚基组成，其作用是固

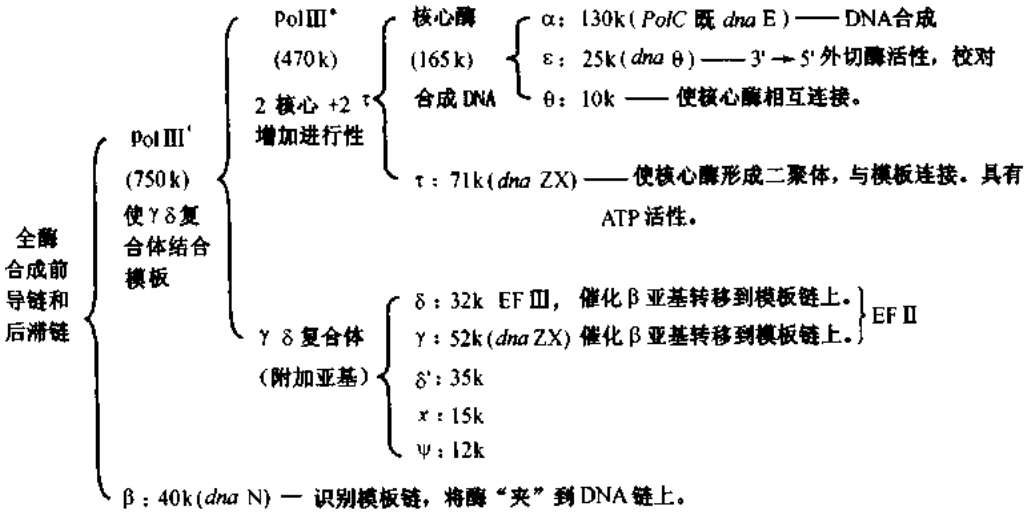


图 11－21 DNA 聚合酶Ⅲ的成分与功能。

定β亚基这个"夹子"的支架。β亚基加盟后就形成 DNA pol Ⅲ全酶。这是一种不完全对称的结构，左侧负责前导链的合成，右侧负责后滞链的合成。

全酶在 DNA 上的装配分为三个阶段：①一个β二聚体加上一个γ复合体识别引物模板形成一种**前起始复合物**。在此反应中，γ复合体水解 ATP 并将β亚基转运到引发模板上。一对β亚基形成一个"夹子"，夹在 DNA 的后滞链上，γ复合体成为这种"夹子的支架"，它负责水解 ATP，移动β亚基结合 DNA；②DNA 在于β，γ复合体结合的位点构象发生改变，而对核心酶产生了高亲和性。使核心酶能与 DNA 结合。③τ二聚体结合核心聚合酶，使其二聚化，也就是与另一个核心聚合酶结合，第二个核心酶又和另一个β"夹子"结合，这样装配好了全酶。值得注意的是全酶只有一个γ复合体。

二、DNA 连接酶

任何聚合酶尽管它们有能力合成长片段 DNA，但却不能将裂缝中相邻的 3′－OH 和 5′－PO_4 连接成磷酸二酯键（图 11－24），只有**连接酶**（DNA ligase）才能做到这一点。其作用机制是分三步进行：

(1)ATP(NAD)　　与酶结合产生 E－AMP＋PPi

$E_{T\triangle}$ + ATP→E－AMP + PPi

在 *E.coli* 中，　　E(酶) + NAD→E－AMP + NMN(烟酰胺单核苷酸)

(2)E－AMP 上的 AMP 转移到 DNA 的 5′－PO_4 上使其活化

(3)活化的 5′－PO_4 与相邻的 3′－OH 作用形成 3′－5′磷酸二酯键(图 11－25)，并释放出 AMP。但连接酶不能连接单链，必须要有模板的存在。也不能连接 3′－OH 和 5′－PPP 以及 DNA 与 RNA。在后滞链的合成，DNA 损伤的修复，重组及转座中 DNA 连接酶都发挥着重要的作用。

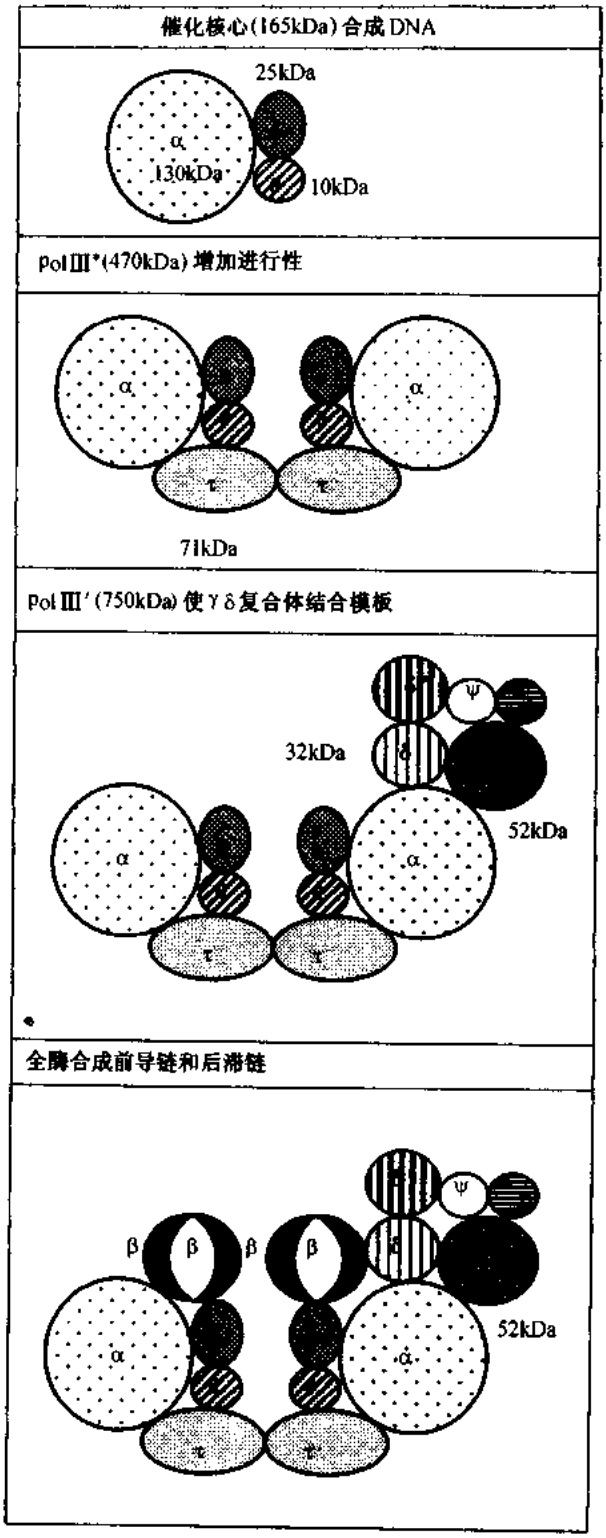


图 11－22　DNA 聚合酶Ⅲ的组成及各部分的功能。(仿 B.Lewin: Genes Ⅵ 1997, Fig.15－12)

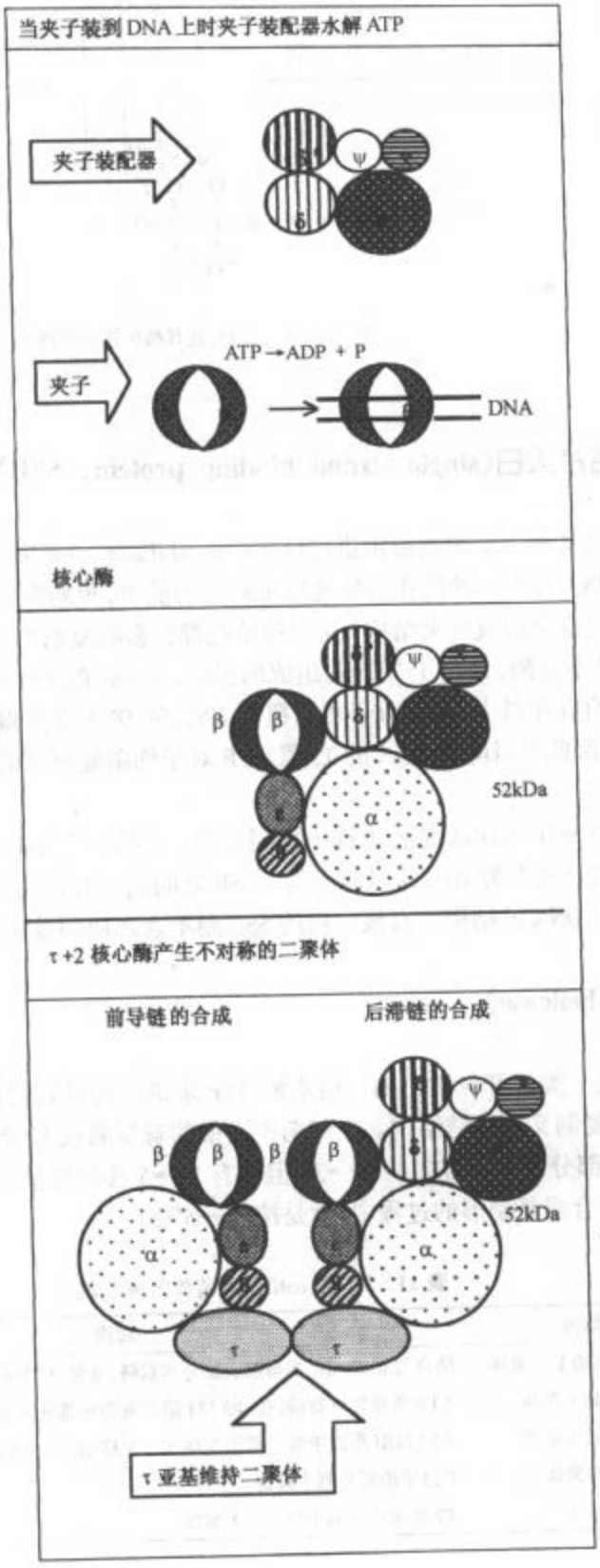


图 11－23　DNA 聚合酶Ⅲ全酶的装配过程。(仿 B. Lewin: Genes Ⅵ 1997, Fig. 15－13)

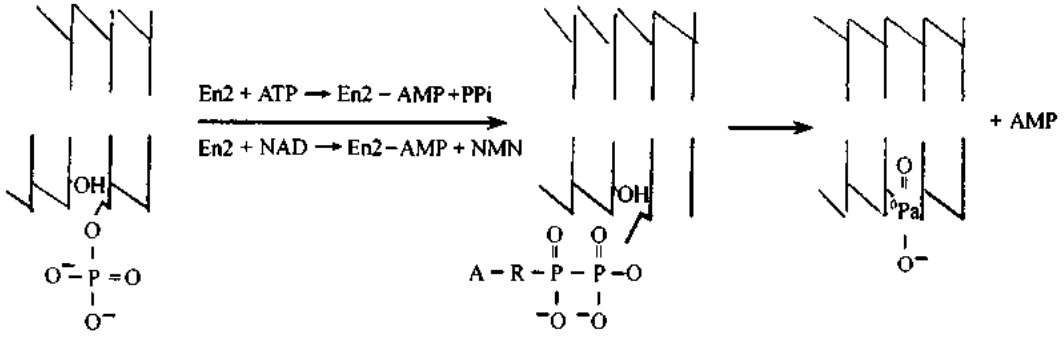


图 11 – 24　DNA 连接酶的作用机理。

三、单链结合蛋白(single strand binding protein, SSB)

这种蛋白又称为双螺旋稳定蛋白(helix destabilizing protein)。当解旋酶将双链打开以后,单链 DNA 具有一种潜在的恢复原来双链的能力,重新形成氢键。而且单链本身若有反向重复也会形成发夹结构,这两种情况都会影响复制的,但 SSB 可以解决这一问题。SSB 并不是酶,是由 177 个 aa 组成的蛋白,*E. coli* 的 SSB 以四聚体存在,分子量为 74kDa,结合在单链上,每个分子可以覆盖 32nt,使 DNA 受到保护,不被酶水解,也不回复成双链,因此可以降低 DNA 的 T_m 值。SSB 对于细菌复制叉的结构的作用和解旋酶是同等的。

在原核中 SSB 与 DNA 结合表现出协同效应。若第一个 SSB 的结合能力为 1,则第二个 SSB 的结合能力为 10^3。这可能因为①SSB 之间的相互作用;②第一个 SSB 和 DNA 的结合改变了 DNA 的结构。真核生物的 SSB 则不表现协同效应。

四、解旋酶(helicase)

解旋酶是一类解开氢键的酶,由水解 ATP 来供给能量它们常常依赖于单链的存在,并能识别复制叉的单链结构。在细菌中类似的解旋酶很多(表 11 – 3),都具有 ATP 酶的活性。大部分的移动方向是 5′→3′,但也有 3′→5′移动的情况,如 n′蛋白在 ΦX174 以正链为模板合成复制形的过程中,就是按 3′→5′移动。

表 11 – 3　*E. coli* 及噬菌体的解旋酶

解旋酶	结构	功能
DnaB	330 k 六聚体	结合与 oriC、oriλ 参与前引发分离双链,水解 ATP 提供能量
rep 蛋白	66 k 单体	ATP 依赖性解旋酶,在 ΦX174 滚环复制中推进复制叉
PriA(w 蛋白)	82 k 单体	ΦX174Rf 形成中参与前引发体 3′→5′移动取代 SSB,识别特异位点
TraY/I	多聚体	F 因子滚环复制中解链
基因 4 蛋白	58 K	T7 噬菌体复制中延 5′→3′解链

第五节　原核生物及噬菌体和病毒的复制起始和终止

前面介绍的一系列实验，已证明大部分原核生物及质粒只有一个固定的复制起始点，那么现在要问这复制点的结构是怎样的呢？为了解决这个问题，人们采用了不同方法进行了研究，那就是逐步酶切，标记获救（marker rescue）；转导噬菌体和 F′因子。将 *E.coli* 的复制起始的区域切下，然后逐步酶切，每次切下的片段都接到一个含有抗药性标记，而没有复制点的质粒上，转化细菌，观察能否复制，一直到被切片段失去复制起动功能为止，然后切下测序。另一种方法是将转导噬菌体整合到复制起点的附近，以复制起始点相邻的基因为标记，然后切除噬菌体，获得带有复制起点的质粒如 pCM 959。

一、环状 DNA 的复制

1. 双链 DNA

(1) *E.coli*

现已知道 *E.coli* 的复制起点（oriC）区长 245bp（+22～+267），结构如图所示（图 11－25）左侧有了 3 个 13bp 近乎相似的回文序列和 4 个 9bp 的重复序列。R1、R3 正向，R2、R4 反向，它们是 DNA 蛋白的结合位点。复制起始区的结构特点是：①富含 A－T，这可能和双链易于解开起始复制有关；②含有多个回文结构（9－14 个 GATC，*E.coli* 含有 11 个）8 个 GATC 较保守，CATC 中的“A”已甲基化。(3)具有 R1、R2、R3、R44 个反向重复序列，作为蛋白结合位点；(4)此序列的右侧毗邻区域有两个起动子，其可能的作用是：①转录产生引物；②产生复制必要的蛋白；③产生调节功能的 RNA；④起转录激活作用。

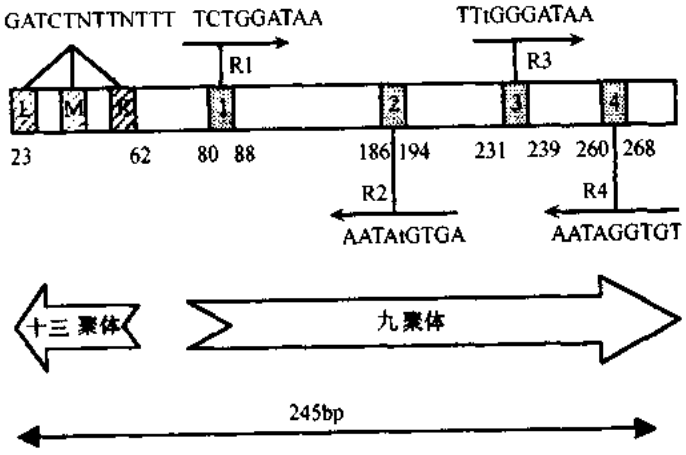


图 11－25　*E.coli* 复制起点 oriC 的结构。

复制的起点已从细菌、酵母、叶绿体和线粒体中鉴定出来，但高等真核生物尚未发现。双链环状 DNA 的复制起始需要一系列的激活作用。①在起始区双链必须打开一

小段，这主要是依赖 RNA 聚合聚在附近的转录，将双链打开，这种作用就叫转录激活（transcriptional activation）；②双链解开的这一点延着 DNA 向边延伸扩大，产生复制叉；③开始合成引物，这个反应在前导链仅发生一次，可是在后滞链上每次形成冈崎片段都要重复这种反应。

在 *E. coli* 的慢停突变中已鉴别出几种蛋白产物和复制起始有关。我们现在已经知道在复制起始早期阶段中 *dna A*、*dna B* 和 *dna C* 的功能，但对 *dna T*、*J*、*K* 了解不多。无论是快停突变还是慢停突变都涉及到 *dnaB* 和 *C*，它们的功能是和起始区的激活以及引发冈崎片段的合成有关。

带有 *E. coli* oriC 的质粒已发展成一种无细胞系统用来研究复制。在 oriC 的复制起始的体外实验表明需要 6 种蛋白，DnaA、DnaB、DnaC、组蛋白样蛋白（HU）旋转酶及单链结合蛋白（SSB）形成起始复合物（表 11－4）。使环状超螺旋模板转变为双链广泛解旋的形式。DnaA 单体首先结合到 oriC 上 4 个含 9bp 的重复序列上。然后 20～40 个 DnaA 单体结合到 oriC 上形成一个核心。在 DnaA 蛋白的作用下位于 oriC 右侧的 3 个含 13 bp 的重复序列开始解链形成开放复合体。这三个 13bp 重复序列的打开对于下一步反应来说是必须的。

表 11－4　在 oriC 上前引发所需的六种蛋白

蛋白	功能	对蛋白质的要求
DnaA(S)	协同结合于 9bp 和 13bp 重复顺序	20～40 单体
DnaB(F,S)	结合于 DnaA，提供解旋酶活性	1～2 六聚体
DnaC(F,S)	和 DnaB 形成复合体	6 个单体
HU	组蛋白样蛋白，激发复合体形成	约 5 个双体
Gyrase	旋转酶，解除正超螺旋，引入负超螺旋	催化
SSB(F)	稳定单链	化学剂量，四聚体

(S)：慢停突变；(F)：快停突变。摘自 B. Lewin：Gene Ⅳ，1990，P353

9 bp 和 13 bp 的重复序列是很独特的，我们还不知道它们是否以相同的或不相同的结合位点被 Dna A 蛋白所识别，但这两种重复序列都是丰含 A·T 碱基对的。

DnaB/DnaC 在 ori 区充当了起始的“引擎”（primosome）。DnaB 和 DnaC 蛋白形成一个复合体。在引发体（primosome）中，每个 DnaB 六聚体结合 6 个 DnaC 单体形成 480kDa，直径为 6nm 的球状聚合体（图 11－26）。在 oriC 处可见到一个大的蛋白质斑点，即是这种复合体。

DnaB·DnaC 复合体转变为 DnaB 六聚物，形成复制叉。DnaB 提供解旋酶（helicase）活性，使 DNA 解旋，可能它识别复制叉上潜在的单链结构，从 13 bp 的重复序列上取代出 DnaA，并开始解螺旋。DnaB 在 ori 区域以很少的量（1～2 六聚物）担负着催化作用。在那儿 DnaB 还具有激活 DnaG 引发酶的能力。

解旋反应还需要另外两种蛋白，旋转酶（Gyrase）和 SSB（单链结合蛋白）。旋转酶也就是 Top Ⅱ（见第 9 章），其作用是解旋，即让一条链绕着另一条链旋转。若没有这步反应，解开双链就会产生 DNA 的扭曲。SSB 可使已形成的单链处于稳定状态。

HU 是 *E. coli* 中的一般结合蛋白（见第 10 章），在体外复制起始并不绝对需要它的存在，但它是能刺激复制起始的。HU 具有结合 DNA 的能力，可能涉及到某些构建的能力。

在不连续复制中，前引发蛋白能在解开 DNA 的一个重要的长度，此长度通常在小

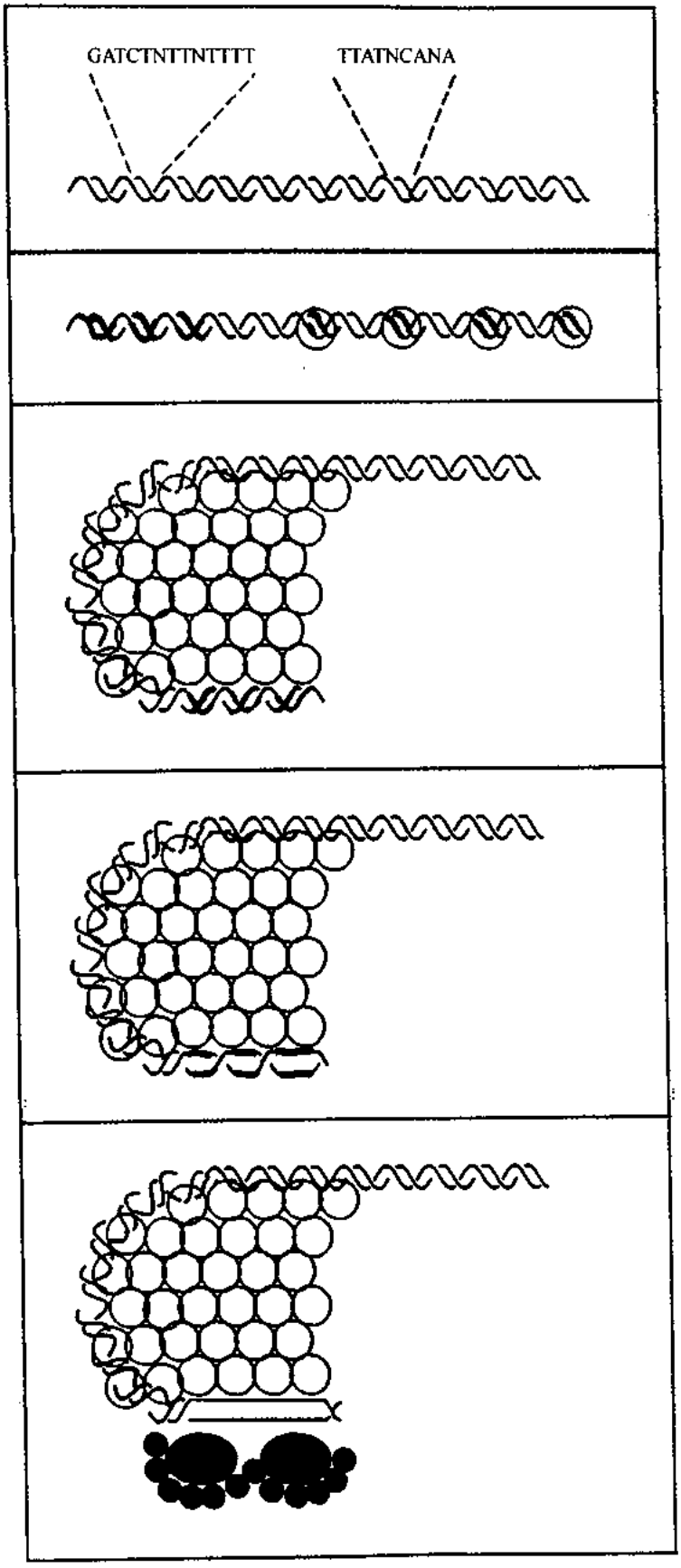


图 11－26 在 oriC 上 DnaA、DnaB 和 DnaC 装配或前引发复合体。（引自 B. Lewin: Genes Ⅵ, Fig. 15－18）

于 60bp 时是不能够起始复制的。

引发反应的能量是从 ATP 中获得的，能量的消耗用于 DNA 的解旋。Dna B 解旋酶的活性是依赖于 ATP 水解，旋转酶的解旋作用也是需要 ATP 水解，引发酶(primase)和 DNA pol Ⅲ的激活都需要 ATP。

复制叉的产生如图 11－28 所示。引发反应的产生一般发生在前导链 (leading strand)上。我们知道：RNA 的合成是起引发作用，但反应的细节并不知道。某些 *dnaA* 突变可被 RNA pol 的突变所抑制，表明 DnaA 可能涉及到体内 RNA 合成起始。RNA pol 可能要经过与转录单位相邻的复制起始区，且在这个区域终止，为 DNA pol Ⅲ提供 3′－OH 的末端引物。转录的作用也可能和改变 DNA 结构、帮助起始有关。这个观点已被证实。在体外转录并不需要进入 ori，它在上游 200bp 远处就可以起作用，而且可以使两条链中的任何一条链作为模板。但在体外对于超螺旋来说就需要经过 ori，表明它是通过 DNA 的结构来起作用的，主要是帮助 DNA 解链。

(2) λ 噬菌体

λ 噬菌体的复制区如图 11－28 所示，共有两个区域。一个是连续的 4 个 O 蛋白结合位点，每个位点 18bp 长；邻接另一个区域是 A－T 丰富区，长 40bp。λ 在 ori 近上游处的复制起始是需要从 P_R 开始的转录链的“激活”，如 oriC 一样，并不需要 RNA 为前导链提供引物。和这个系统相似，表明 RNA 的合成可能涉及到起动起始区某些结构的改变。

λ 噬菌体 O 和 P 基因的产物是复制起始所必须的。O 蛋白结合到 λ 的 ori，而 P 蛋白和 O 蛋白相互作用。还和细菌的蛋白相互作用。ori 区域位于 *O* 基因中，因此 O 蛋白的作用紧靠其合成位点。

λ 的变异体称作 λdv，它含有较短的基因组，携带所有的复制所需的信息，但失去了感染的功能；在 *E.coli* 中得以存活的 λdv DNA 是作为一种质粒。这种 λdv DNA 也能通过一个含有噬菌体编码的 O 和 P 蛋白以及细菌复制的系统进行体外复制。

λ 蛋白 O 和 P 形成一个复合物在 λ 的 oriλ 上和 DnaB 结合在一起。

起始的第一步是 O 结合成一个粗糙的直径为 11nm 的结构，称为 O 体(O－some)，此 O 体含约 100bp DNA(这段 DNA 的分子量为 60kDa)，其中有 4 个长 18bp 的 O 蛋白结合位点，分子量为 34kDa。每个位点都有一个回文序列(TCCC－GGGA)，这样就能结合对称的 O 二聚体。但每个位点的序列不完全相同，当 O 的量较少时，优先和Ⅱ、Ⅲ结合。O 结合位点的序列呈弯曲状，当结合 O 蛋白后便进一步弯曲(图 11－30)。

若 DNA 是超螺旋的话，O 蛋白的结合会导致 ori 区结构的改变。A－T 丰富区就连接在 O－结合位点的边上，成为 S1 酶的敏感位点。S1 酶是特异性识别单链 DNA。表明解链反应就发生在 O 蛋白复合体的边上。

O 蛋白的作用是和 oriC 上的 DnaA 相似。它为 DnaB 的结合作为准备。λ 具有特有的 P 蛋白，相当于 DnaC 的功能，导致 DnaB 和 ori 区结合。当 λ 的 P 蛋白和细菌的 DnaB 结合形成一个大而对称的复合体，它含有较多的 DNA(总量为 160bp)。λ 的 P 蛋白有一个特殊的作用，就是可以抑制 DnaB 的解旋酶活性。P 从复合体中释放出来时就触动了复制叉的移动，随之引发复制，DNA 开始合成。

有的蛋白对复制来说是必要的，但没有直接涉及 DNA 的合成。一个有趣的例子就是 DnaK 和 DnaJ 蛋白。DnaK 是一个分子伴侣(chaperone)，它和真核的应激蛋白(stress protein)相关。它和其他蛋白相互作用，以构象依赖(conformation dependent)方式在很多细胞活性方面发挥作用。DnaK/ DnaJ 的作用可能是分解前引发复合体 (prepriming complex)，导致 P 的释放，随之开始复制。表 11－5 已表示了在 oriC 和 oriλ 在起始复制

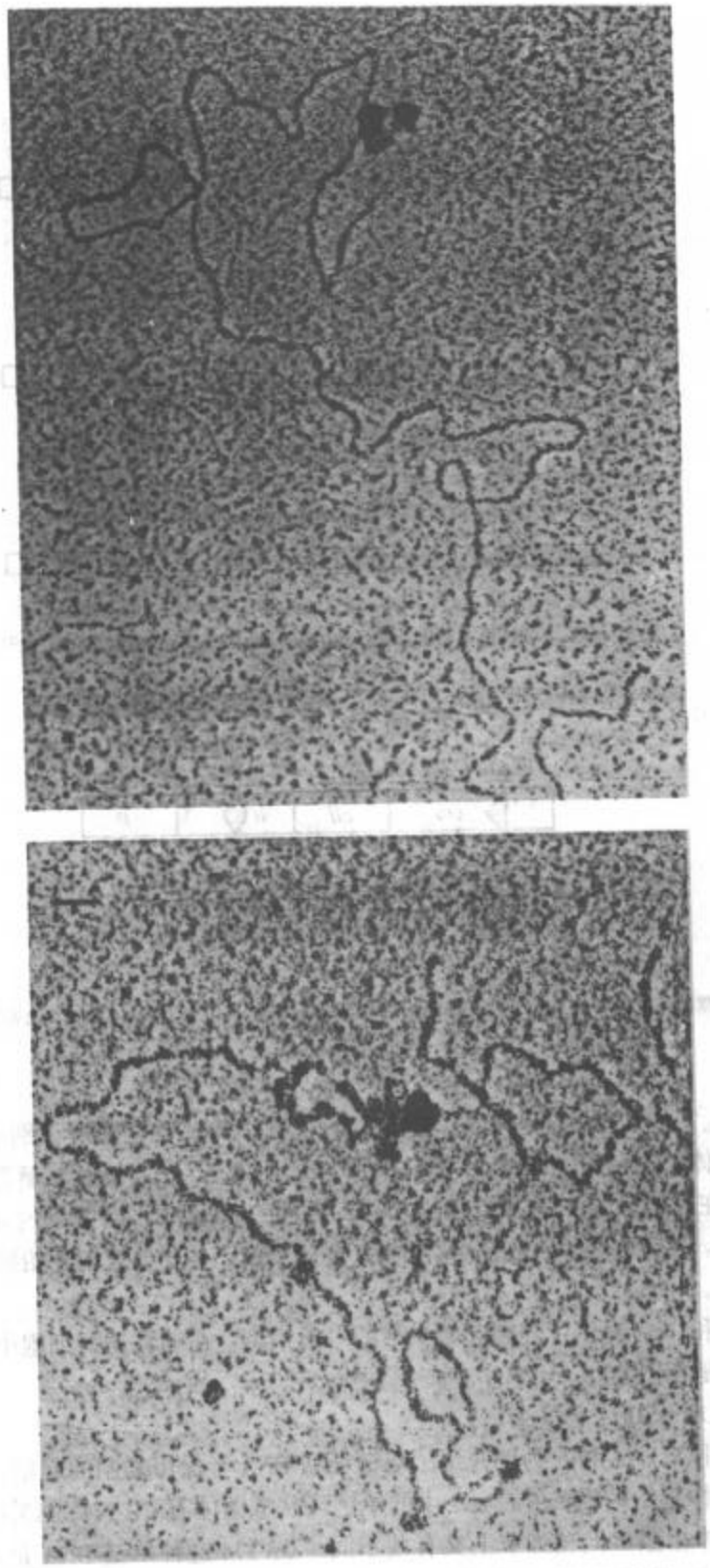


图 11－27　通过电镜可在 oriC 确定是否存在复合体。上图表示开始复制前的前引发复合体。下图表示开始复制后 1 分钟时在复制泡上的复制体。被观察到的复合体和 DnaB 蛋白的抗体结合所显示出的。(转引自 B. Lewin: Genes Ⅵ 1997, Fig. 15－19)

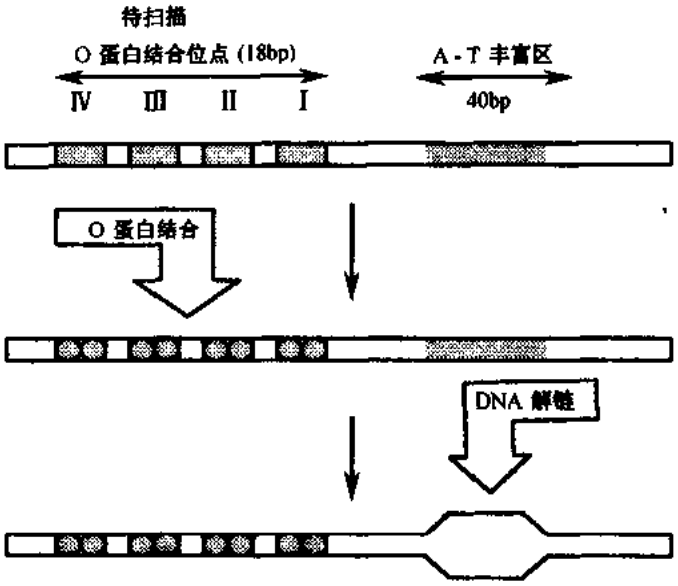


图 11 - 28　λ噬菌体复制的起始。(仿 B.Lewin:Genes Ⅵ 1997,Fig.19 - 19)

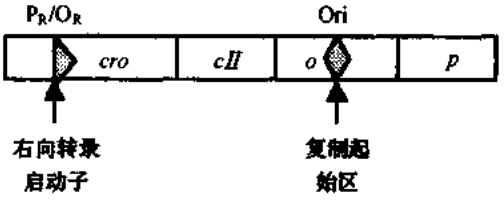


图 11 - 29　λ噬菌体右向转录启动子对复制起始的激活是必须的。(仿 B.Lewin:Genes Ⅵ 1997,Fig.19 - 18)

中的相似性。它们涉及的阶段相同,依赖重复的元件。第一步是蛋白识别 ori,和 DNA 结合形成复合体。DnaA 识别 oriC,O 蛋白识别 oriλ,短的 A - T 丰富区被解链。然后 DnaB 被装配上,在 oriC 和 oriλ 上需要不同功能的蛋白,在此阶段,对于别的 ori 区,还需要其他的蛋白。当解旋酶加入到复合体中时,复制叉就产生了。复制开始后 RNA 的引物最终产生了。

用 oriC 和 oriλ 提供了复制起始激活的一般模型。在哺乳动物细胞中的病毒 SV40 复制起始也是相似的。

(3)单链环状 DNA 的复制

前面已用 *E.coli* 和 λ 为例介绍了双链环状 DNA 半保留复制的起始,那么单链环状 DNA 又怎样起始的呢?在 *E.coli* 中的单链环状噬菌体有 *ΦX*174,M13,G 等。它们的复制类型大体相似,但在细节上又有些不同。*ΦX*174 这种单链 DNA 是 Sinsheimer 于 1959 年发现的,是单链复制系统中最为复杂,也是研究得最为透彻的噬菌体。单链环状 DNA 的复制一般都要经历 3 个阶段,以 *ΦX*174 为例(如表 11 - 6 所示)。第一阶段以亲本链(+链)为模板合成互补的环状负链,形成闭合环状的复制形(RF1)。第二阶段以成环滚环复制(looped rolling circle replication)产生多个子代 RF;第三阶段以 RF 的

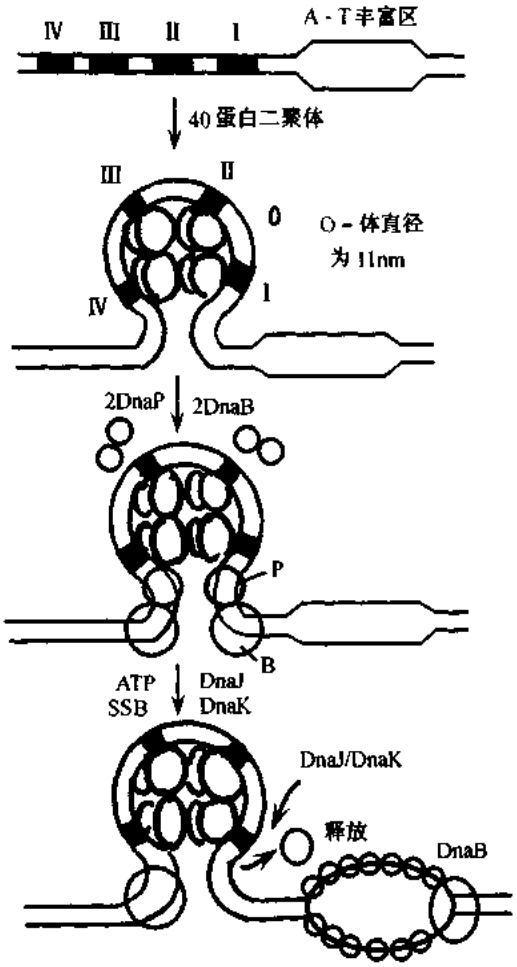


图 11-30 λ 的复制起始。

表 11-5 在 oriC 和 ori λ 上起始涉及相同的反应

阶段	oriC	ori λ
ori 的识别与解链	DnaA	O
解旋酶的结合与解链	DnaB	DnaB
	DnaC	P
	RNA pol	RNA pol
释放复合体	DnaJ?	DnaJ?
	DnaK?	DnaK
复制叉的移动	Gyrase	?
	SSB	?
引发	DnaG	DnaG

摘自 B. Lewin: Gene V, 1994, Table 19.6

负链为模板进行滚环复制产生多拷贝正链单环(图 11-31)。

*ΦX*174 的第一阶段复制中又可分为五步。当进入寄主后单链结合蛋白(SSB)结合在单链环上(正链),但在 *F* 和 *G* 基因之间有一发夹结构,长约 55nt,不被 SSB 结合,成为裸露的双链,可与 5 种蛋白(PriA、PriB、DnaT、DnaB、DnaC)结合成前引发体

(preprimosome)。这个位点称为引发体结合位点(primosome assembly site,pas),也是惟一的引发体结合位点。Pri A 识别 pas。M13 有 6407 个 nt,在 5480 ~ 5820nt 之间有 5 个发夹结构,其中第 3 个最重要,有 59 个 nt。*E. coli* 的聚合酶识别此发夹结构形成复制起始复合物。G4 有 5577 个 nt,也是在 F 和 G 基因之间有三个发夹结构。引物就结合在发夹结构处，合成引物,不需形成前引发体(图 11 - 32)。

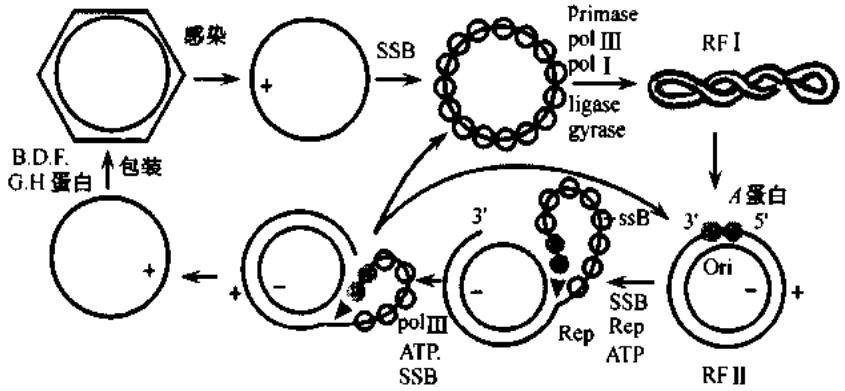


图 11 - 31　*ΦX*174 DNA 合成周期。

表 11 - 6　*ΦX*174 的复制周期

阶段	时间(分)(在 33℃)	行为
SS→RF	0 ~ 1	吸附和穿透。病毒单链经复制成复制形(RF Ⅰ),RF Ⅰ转录
RF→SS	1 ~ 20 25	亲代 RF 经复制产生 60 个子代 RF,RF 和宿主 DNA 的半保留复制停止
RF→SS	20 ~ 33 40	子代 RF 形成 35 个滚环,复制出 500 个 ss 分子,包装入噬菌体颗粒,细菌裂解

David Freifelder: Molecular Biology, 1983, Table 15 - 8

在 *ΦX*174 的前引发体中，DnaB 六聚体和 6 个 DnaC 单体结合形成复合体，在 ATP 存在下对单链有很高的亲和性。并能使 DNA 结构发生改变，为引发酶合成引物 RNA 创造条件(图 11 - 32)。引发酶的加盟,和前引发体形成引发体,完成了"组装"的任务。第三步是移动，引发体是在 *ΦX*174 的 pas 位点上形成的，但引物合成是在很多位点上,这表明引发体会沿着单链 DNA 引发位点移动(图 11 - 32)。当引物合成后,PriA 水解 ATP,DnaB 是变构,降低了与单链的亲和性,引发体至少含 DnaB,PriA 和引物酶,还不清楚是否有别的蛋白。PriA 不仅能识别 pas 位点,还具有水解 ATP 和置换 SSB 的双重活性,这对于引发体可以移动到其他位点进行引发是十分重要的。引发体移动的停止位点的性质也不清楚。在 M13 和 G4 中和 *ΦX*174 不同引发点仅有一个。第五步延伸,则是依靠 DNA pol Ⅲ在引物的 3′ - OH 不断加入新的核苷酸，使链延伸直至前面引物的 5′端,形成多个冈崎片段,再通过 DNA pol Ⅰ切除引物,从后面的片段 3′起始合成短片段的 DNA,将空缺补起来,最后留下的裂缺由连接酶将其封闭,从而完成复制,形成了复制型Ⅰ(RF Ⅰ)。带有缺刻的双链环称之为 RF Ⅱ。

*ΦX*174 复制的第二阶段是由成环滚环复制和负链的不连续合成来完成的,第三阶段也是成环滚环复制。复制起始区的结构如图 11 - 33 所示。

滚环复制又叫 σ 复制(图 11 - 34),和 θ 型复制的模式不同,θ 复制是全程起始(de

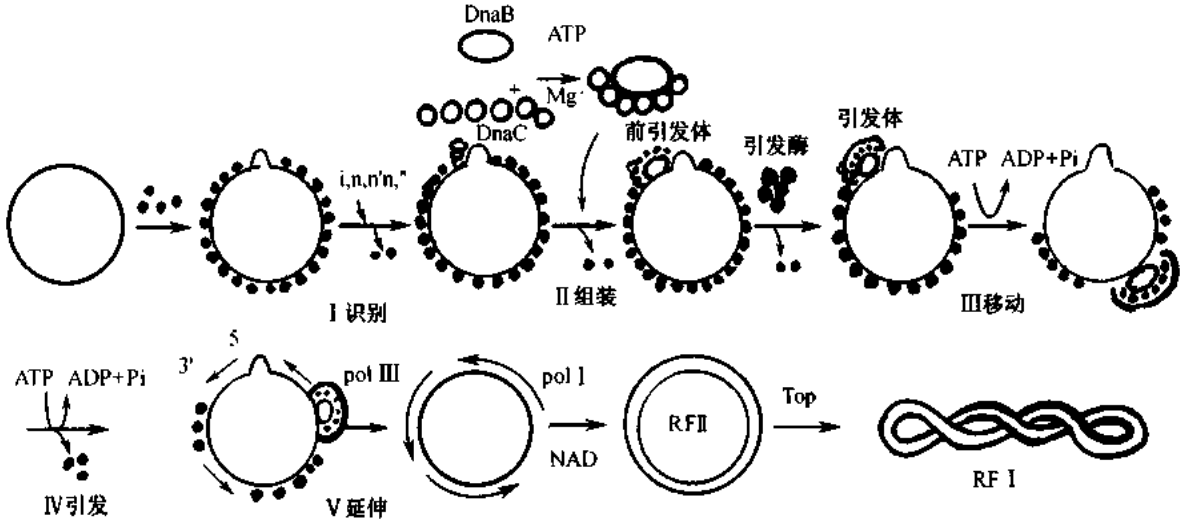


图 11－32 ΦX174 第一阶段复制形成 RF Ⅰ。

novo initiation)，而 σ 复制是共价延伸，1968 年由 Gilbert 提出的，它有以下的特点：①共价延伸。θ 复制是以引物的 3′－OH 为起点以亲代链为模板从头合成一条新链，新链和旧链是完全分开的。但 σ 复制是先在一条亲代链(＋链)上切一缺口，然后以另一负环为模板，在正链切口上的 3′－OH 上延伸，新合成的链和正链连在一道；②模板链和新合成的链分开。θ 复制是半保留的新合成的，一条单链总是和模板链互补地结合在一起形成一条子链。而 σ 复制 3′ 端不断延环状模板合成，5′ 端脱离负环，随着复制游离的 5′ 端越来越长，和环形模板完全分开；③不需 RNA 引物，在正链 3′－OH 上延伸；

表 11－7 *E.coli* 中与复制有关的基因

基因	产物	功能
dnaA	DnaA 52kDa 单体	20－40 单体结合到 oriC 的 9bp 和 13bp 的重复序列上
dnaB	DnaB 330kDa 六聚体	解旋酶(5′→3′)，参与前引发，分离双链，在单链上用于水解 ATP，提供能量
dnaC	DnaC 29kDa 单体	六个单体，分别结合于 DnaB 的六聚体上，组成引发体
dnaT	i 蛋白 66kDa 三聚体	参与前引发，(1－2 分子)/每个复制叉
dnaG	引发酶 60－kDa 单体	起始冈崎片段，合成 RNA
dnaY	$tRNA^{Arg}$ 118bp	不知
dnaE	DNA pol Ⅲα，130kDa 亚基	DNA 聚合和校对
dnaZX	DNA pol Ⅲγ，52kDa 亚基	不知
dnaZX	DNA pol Ⅲτ，71kDa 亚基	与模板结合，在蛋白水解酶作用下产生 γ 亚基
dnaQ	DNA pol Ⅲε，27.5kDa 亚基	3′→5′校对功能
dnaN	DNA pol Ⅲβ，40.5kDa 亚基	链的延伸
lig	连接酶，75kDa 单体	封闭冈崎片段之间的裂隙
ssb	SSB，76kDa 四聚体	单链结合蛋白，使其保持稳定，防止形成双链
PriA	n′82kDa 单体	位点识别，解旋酶(3′→5′)(一个分子/复制叉)取代 SSB
PriB	n″23.5kDa 单体	参与前引发
PriC	n24kDa 单体	参与前引发
rep	解旋酶 66kDa 单体	ATP 依赖性解旋酶，复制叉的移动

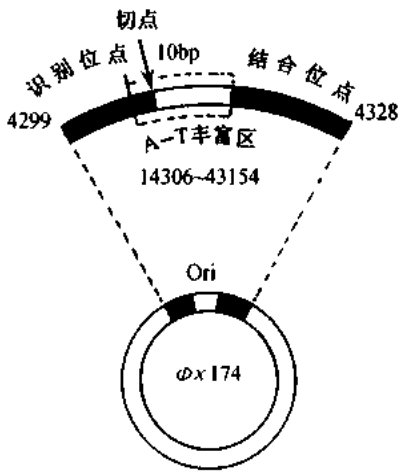


图 11－33　$\Phi X174$ 滚环复制起始区。

环复制的过程相似，所不同的是 5′ 端游离部分，不再复制而单独形成环，每环含一个基因组。$\Phi X174$ 滚环复制的起始区含 30 nt 的保守顺序，构成了 A 蛋白的识别位点、结合位点和 A－T 丰富区（图 11－33）。A 蛋白由 $\Phi X174$ 自己编码，在第一阶段复制完以后转录翻译而产生，也是复制体系唯一由 $\Phi X174$ 本身编码的蛋白，其余的蛋白都由 *E.coli* 编码。A 识别 Ori 的特殊顺序后，诱导了 A－T 丰富区的解链，并在（＋）链的第 4305～4306nt 之间切断磷酸二酯键，产生 3′－OH 和 5′－P，A 蛋白的 Tye 的－OH 与 5′端的 nt(A)形成磷酸脂链，结合于其上。这个反应要求 DNA 呈负超螺旋状态。这样 A 蛋白能结合在含 10 个 nt 的单链片段上，在其一端切开（＋）链，产生裂缝。也就是说负超螺旋为 A 蛋白提供单链结合位点是十分必要的。

F 因子的转移是在 oriT 位点起始的。它位于起始区的一端（图 11－36），TraY 和 TraI 在 oriT 的一条链上切一缺口，然后这些蛋白作为一个大的复合体结合到 DNA 上，导致 200bp DNA 的解链。TraY/TraI 多体沿着 DNA 从 5′端移动，以每秒 1200bp 的速度解旋，图 11－36 表明 5′游离端被

④只有一个复制叉。θ 复制一般都有两个复制叉。⑤形成多联体(concatemer)。5′游离的部分经半保留复制形成双链，并含有多个基因组，串联在一起称为多联体。λ 的增殖、接合，以及真核生物 rDNA 的扩增都是以这种形式。

参与滚环复制的除 A 蛋白和 SSB 以外，还有 Rep(解旋)、DNA pol Ⅲ、DNA pol Ⅰ、RNA pol、连接酶等，形成复制体(图 11－34)，催化连续复制和半不连续复制。成环滚环复制(looped rolling circle replication)基本上和滚

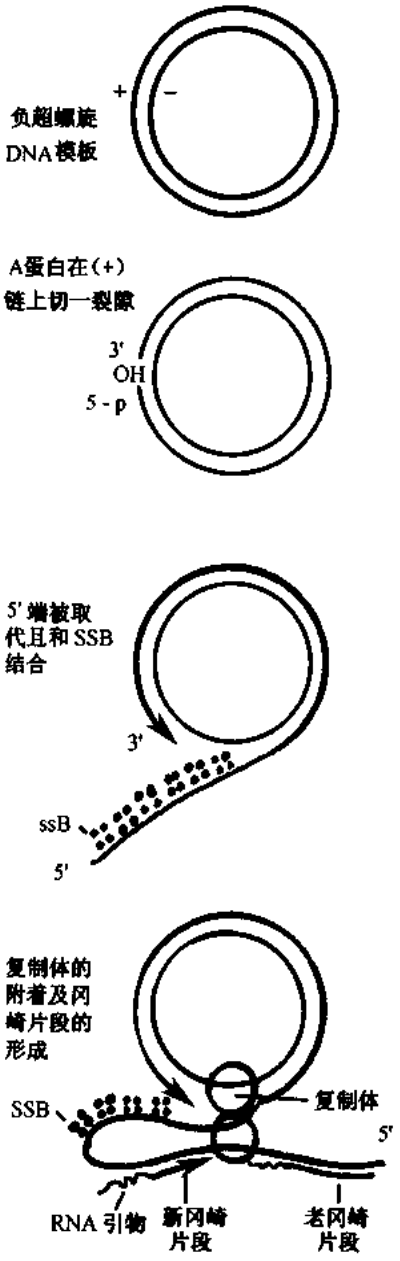


图 11－34　滚环复制的复制体模型。

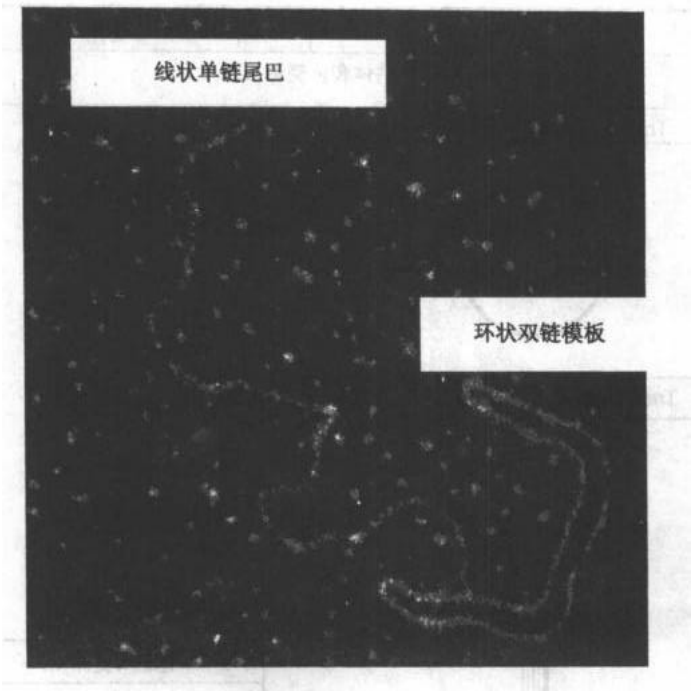


图 11－35 滚环复制的电镜照片。滚环是以一种带有线状尾巴的环存在的。
(转引自 B.Lewin: Genes Ⅵ, 1997, Fig. 14－17)

导入受体菌中，在受体菌中合成此单链的互补链，是细菌呈 F^+状态。供体中余下的一条环状单链也必须互补合成为双链。若此发生在 DNA 转移的过程中，F 质粒的状态将类似于滚环复制。不产生延伸单链区（图 11－36）。通常接合状态的 DNA 像是一个滚环，然而转移链被滚环机制所取代，这样的复制不需提供能量，而单链的转移是DNA 的自由合成。我们还不清楚这样一个特点，当滚环复制时，只有一个单位长度的 F 因子转移到受体细菌中。

二、线状 DNA 的复制

线状 DNA 的复制的起始有二种类型，一种是从中间起始，如 T7，另一种是从末端起始，如腺病毒。线性复制有个发人深思的问题：已知所有的聚合酶都是从 5′→3′方向合成，而且必需在 3′－OH 上延伸，自己本身不能起始，那么线性复制中的从末端起始，其引物从何而来？

1. 双链线状 DNA 的复制起始

(1)T7 噬菌体的复制起始

T7 是线状 DNA 噬菌体，整个基因组有 39,936 bp，在标明的 41 个基因中(图 11－37)，已鉴定了 34 个基因的产物，确定了 26 个基因的功能。

T7 的复制起始区(图 11－38)位于 1～1.1 之间，在 P Ⅱ处。在基因组的近左端约 17%处，有长 167 bp 的区域，其中包括两个长 23 bp 的启动子，序列不完全相同，但同

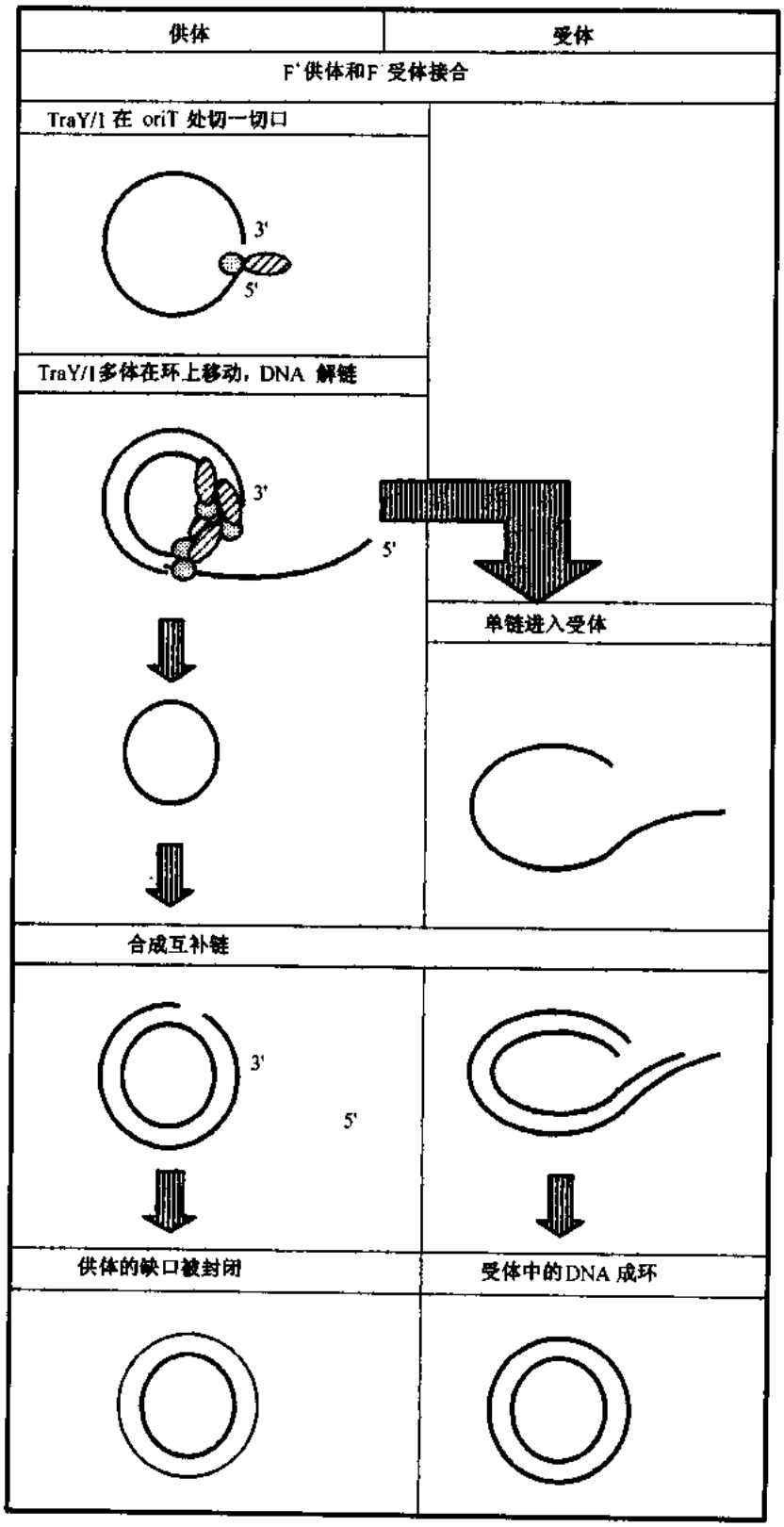


图 11－36　F因子 DNA 的转移。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 14－22)

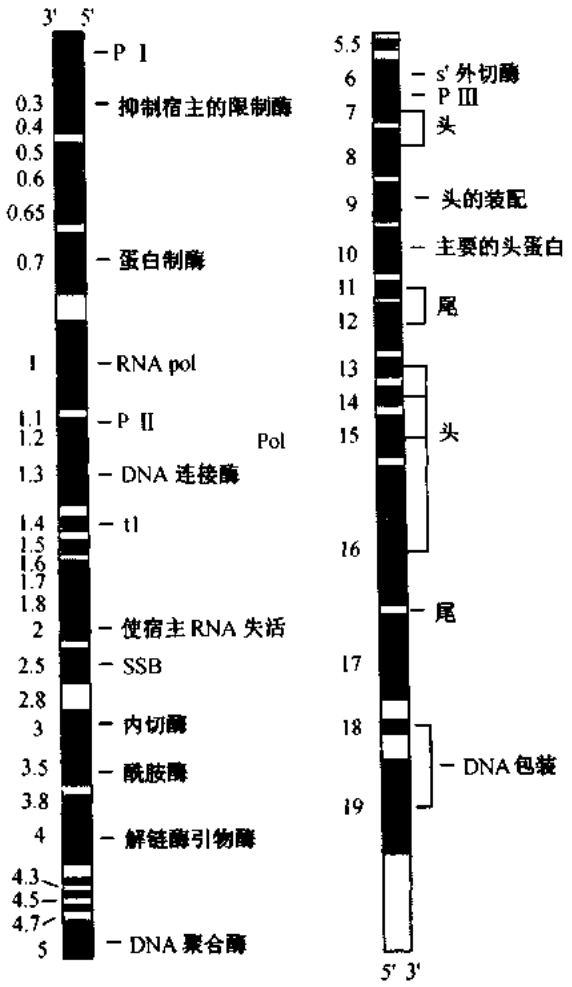


图 11－37　T7 噬菌体的遗传用谱。起动子：P Ⅰ，Ⅱ，Ⅲ，终止子：t'

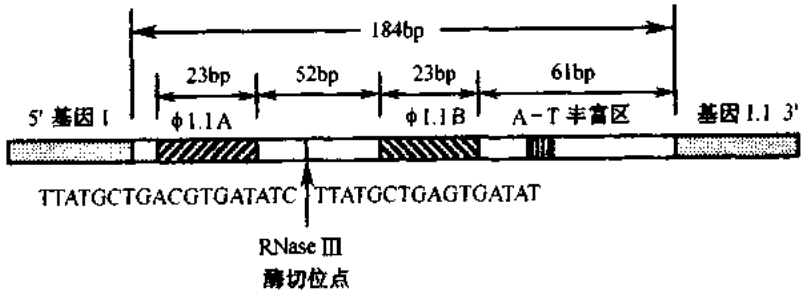


图 11－38　T7 噬菌体复制起始区的结构。

源性很高。另外还邻接一个长 61 bp 的 A－T 丰富区。在两个启动子之间有一 RNase Ⅲ的切点；在 A－T 丰富区中有 7 个短的回文序列 TTAA。

T7 复制所涉及的酶和蛋白如表 11－8 所示。复制时 ϕ1.1 A 和 ϕ1.1 B 两个启动子中任一个起动都可以起到转录激活的作用，在邻接的 A－T 丰富区打开双链。并为前导链的合成提供了引物。在 A－T 丰富区 DNA pol 利用 RNA 3'－OH 为起点继续合成 DNA 前导链。A－T 区形成开放复合体后，基因 *4*　58kDa 的产物就结合在一条单链上，水解三磷酸核苷，为复制提供能量，并起到解旋酶活性，沿 5'→3'方向解开双链。在 A－

表 11－8　T7 DNA 合成所需的酶和蛋白

酶	基因	结构和功能
RNA pol	*1*	98.1kDa 单链，可能合成引物
DNA pol	*5*	84kDa，ssDNA3′→5′外切活性，有聚合酶活性，组成 DNA pol 亚基
	trXA	宿主基因合成 12kDa，硫氧还蛋白(thioredoxin)，pol 亚基
解旋酶	*4*	58kDa 的产物
二磷酸酯酶	*4*	58kDa 的产物
引发酶	*4*	66kDa 的产物，识别 5′GGGTC3′或 5′TGGTC3′
SSB	*2.5*	单链结合蛋白
内切酶	*3*	降解宿主 DNA 用于合成
5′外切酶	*6*	同上
连接酶	*1.3*	封闭缺口

T 丰富区遇到 5′ TGGTC 3′时，激发引发酶活性，合成短片段 pppACCC 或 pppACCA 的 RNA 作为后滞链的引物。

(2)腺病毒复制的起始

线性双链 DNA 的复制有的是从一端开始的，像这种情况引物如何提供呢，这是一个十分有趣的问题，现已分离了一些与末端起始复制有关的蛋白。研究得较清楚的是腺病毒（adenovirus），ϕ29 DNAs 和脊髓灰质病毒（poliovirus），现以腺病毒为例来介绍 DNA 末端起始复制。

腺病毒全长 35937 bp，是一种较大的线性 dsDNA，两端各有反向重复序列 103～162 bp，两末端各有 50 bp 为复制起点。其复制是以单链置换(strand displacement)的机制来解释的。这样的起始也可以发生在另一端。新链的合成开始于一端，取代了同源非模板链。当复制叉到达分子另一末端时，被置换链释放出来。形成一条游离的单链。它的复制是独立完成的，首先是通过两端的反向重复序列相互配对，形成一个大的茎环，或叫平锅形结构(图 11－39)。

有些病毒被发现有一种蛋白附着在两端的 5′ 端。在腺病毒中是由 55kDa 的蛋白以磷酸二酯键的形式连接在成熟病毒 DNA 上(图 11－40)。在 ϕ29 中也是如此。在单链 RNA 的脊髓灰质炎病毒中是仅有 22aa 的 VPg 蛋白连接在 5′ Tyr 端的羟基上。各种类似机制的附着蛋白都是由病毒编码的，并在附着中发挥重要作用。在腺病毒 DNA 的复制链上的蛋白是 80kDa，称为末端蛋白(TP)，它比附着在成熟 DNA 上的要大。这种 80kDa 蛋白实际上是起始复制的特异蛋白。但在成熟的病毒中，它被切成较小的蛋白保留在病毒中。

在腺病毒感染细胞的抽提物中，80k 的腺病毒蛋白能共价连接 dCTP。这个蛋白和另一蛋白——140kDa DNA 聚合酶紧密连接(图 11－41)。80kDa 的蛋白和 DNA 聚合酶形成一个复合物。80k 末端蛋白丝氨酸残基连接的胞苷酸处在腺病毒 DNA 的末端，其游离的 3′－OH 为 DNA pol 延伸反应提供引物。产生了从 CTP 起始的新链。在 ϕ29 的复制中可能存在同样的情况。整个的反应系列是协调的。80kDa 蛋白对 dCTP 的连接是由腺病毒 DNA 上的 DNA pol 来承担的，此表明连接反应和合成 Nt－Nt 链的方法相同。80kDa 结合在距 DNA 末端 9 和 18bp 之间的区域。此“核心顺序”在宿主的核 DNA 中也有所发现。相邻的 17 和 48bp 之间的区域是宿主蛋白核因子 I 结合所需的位置，核因

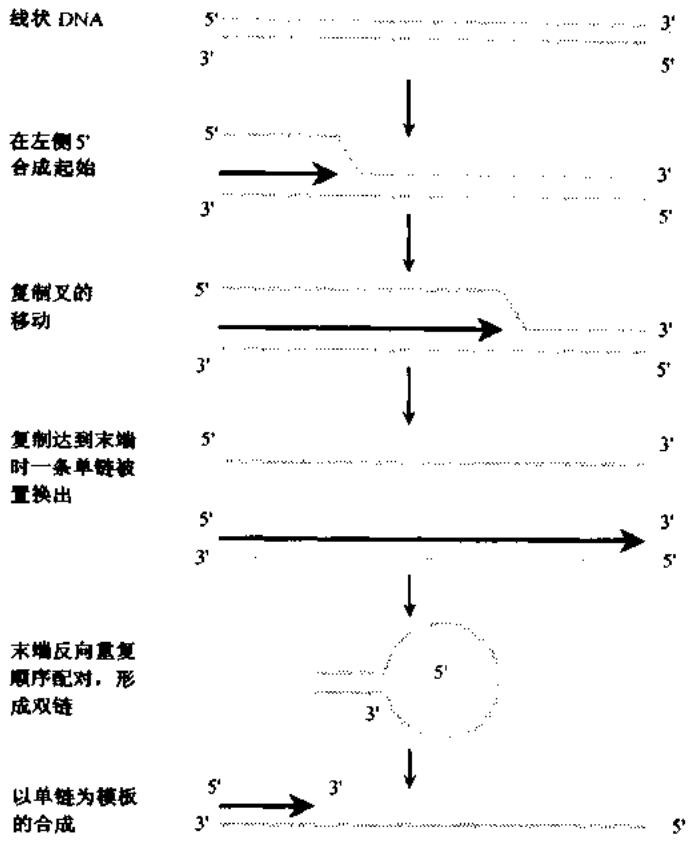


图 11－39　腺病毒 DNA 复制的单链取代模型。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 14－13)

子 I 也是复制起始所必需。起始复合物可能在 9～48 bp 之间的位置，离 DNA 的实际末端有一定的距离。由复制产生的这一结构依赖于模板和复制叉之间的关系。无论是环状的模板（滚环复制）还是线状模板，无论在合成中复制叉占有两条链还是一条链，这种结构是一个重要的特点。

2. 单链线状病毒的复制

啮齿动物中的一种病毒叫微小病毒(parvovirus)，它的基因组是一条单链线状 DNA，一个很有趣的问题就是这种单链线状 DNA 的复制如何起始?引物如何提供?

微小病毒一般含 4800bp，仅编码三个蛋白，具有重叠基因。这三种蛋白都是用来建造外壳蛋白的，转录和复制所需的蛋白和酶系统都是利用宿主的，因此与宿主 DNA 同步复制，但复制的模式十分特殊。

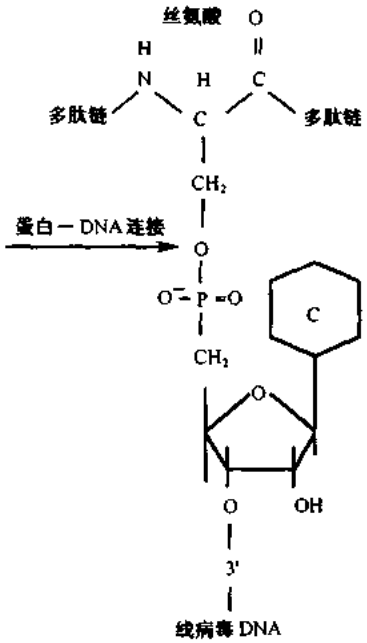


图 11－40　线病毒 DNA 和 55kDa 蛋白的丝氨酸连接。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 14－14)

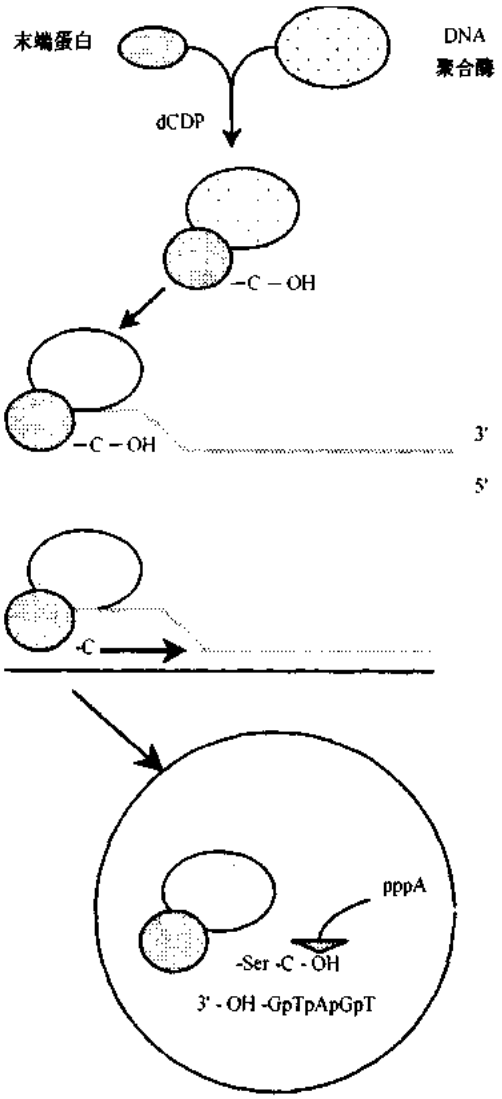


图 11 - 41　线病毒末端蛋白结合在 DNA 的 5′端并提供一个 C - OH，作为新合成 DNA 的引物。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 14 - 15）

微小蛋白单链 DNA 的两端具有不同的反向重复顺序（图 11 - 42 Ⅰ），可以形成发夹结构，其中一端具有 3′ - OH 末端，为复制提供了“引物”。DNA pol 就以单链为模板，在 3′ - OH 起始合成，不断延伸，复制完成后，形成了一个完整的双链发夹结构（图 11 - 42 Ⅱ）。然后再在模板链 3′端反向重复连接处切开单链（图 11 - 42 Ⅱ以箭头所示），再从切开的 3′ - OH 延伸，形成了完整的线性双链分子（图 11 - 42 Ⅲ）。这种双链线性分子的两端再形成发夹结构，为下一轮复制提供了“引物”。这样双链都可作为模板进行复制，也可只用其中一条作为模板。

三、复制的终止

复制的终止主要涉及两个问题，一是环状 DNA 的复制终止位点有没有特殊的结构，它是如何终止的？二是在半保留复制中的后滞链中，最初的冈崎片段上的引物要被切除，这样会形成一个空缺，在环状 DNA 复制中可由与其相邻的前导链提供 3′ - OH，由 pol Ⅰ合成短片段 DNA 来填补这个空缺，而像 T7 这样从中间开始复制的线状 DNA

双链的 5′ 空缺是如何填补的呢?

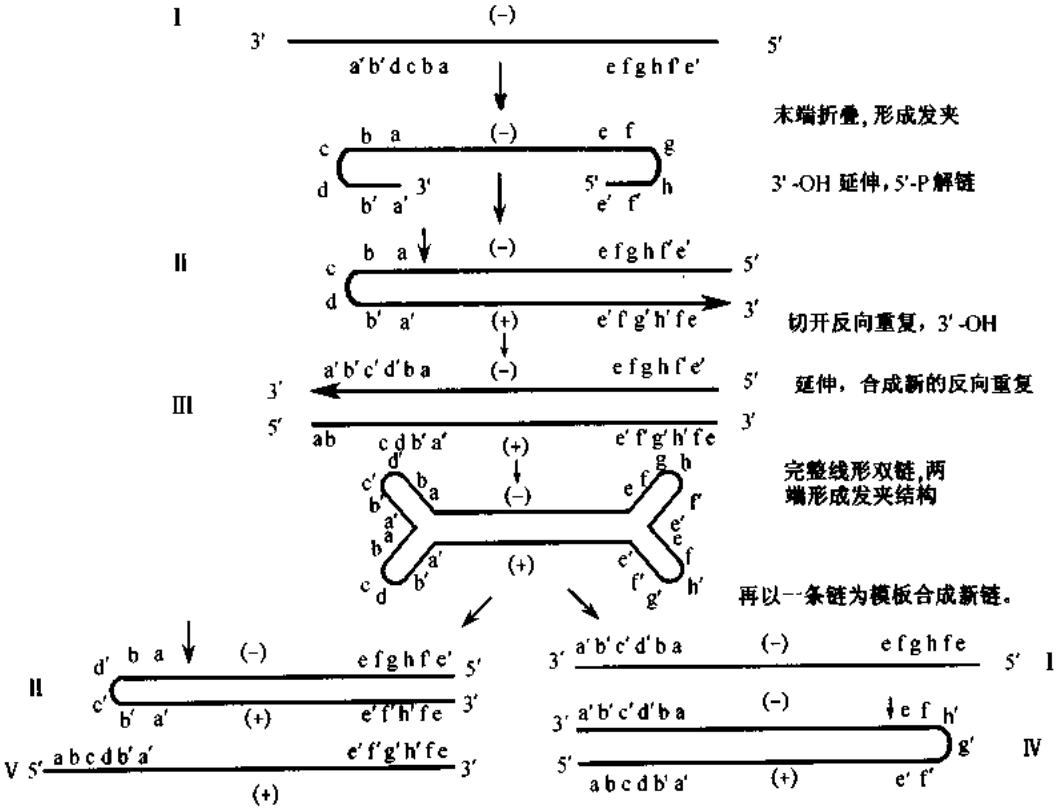


图 11-42 微小病毒 DNA 的复制。(仿 D.Freifelder: Molecular Biology, 1983, Fig. 21-33)

1. 环状 DNA 复制的终止

对细菌环状染色体来说,两个复制叉反向而行,在终止位置相遇。它们怎样相遇呢?是相互碰撞,还是有一个特殊的终止位置呢? *E.coli* 的两个复制叉是要穿越相遇处彼此交叉一直到达特殊的终止位点(图 11-43)。*E.cloi* 的终止顺序是很有特色的。人们现在已经知道复制叉通常会相遇的,但也会随着染色体进行到半途中停下来。现已发现有两个终止区域(*ter D*, *A* 和 *ter C*, *B*),位于相遇点的另一侧 100kb 处。每一终止序列对某一方向移动的复制叉来说是特异的。按这种排列使得每个复制叉要想到达它本身所敏感的终点就必须穿越另一个复制叉的终点。这种排列形成一种“复制叉陷阱”;若由于某种原因一个复制叉移动被延误了,那么就不能在“老地方”会面了,移动较快的那个复制叉将会“陷”在 *ter* 区,等待另一个复制叉的到来。

ter 顺序有一个 23bp 的区域,在体外可导致复制的终止,但其功能显示了一定的方向性。与此相同的顺序在一些质粒中也有发现。在一般情况下复制终止是发生在靠近 *ter* 保守序列处,现在还需进一步弄清楚 *ter* 序列和复制实际停止位点之间的关系。终止需要 *tns* 基因的产物,*tns* 基因编码的蛋白能识别 *ter* 保守顺序,并阻止复制叉继续前进。

有的环状染色体并没有特定的终止位点,如 λ 噬菌体就是通过简单的碰撞而终止的。人们研究了 DNA 大量缺失的 λ 突变株,缺失片段并不一定在复制起点的对面,若 ori 对面有终止位点的话,这个终止位点无论缺失在其他什么地方,都会使终止位点发生偏离,一侧离 ori 较近,另一侧离 ori 较远。但据观察这些突变株的两个复制叉依然双向对称移动,终止于 ori 的对方,可见并不存在特定的终止位点。

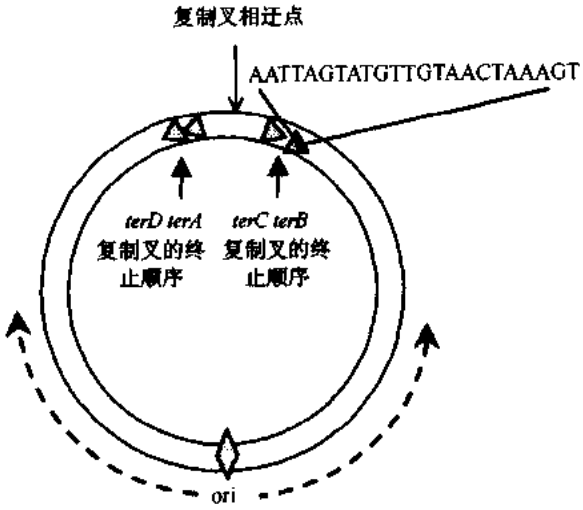


图 11－43　*E. coli* 的复制终止位点。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig. 14－7)

2. 线状 DNA 分子的复制终止

按半保留复制的双链 DNA 理论说，当 5′ 端引物被切除后，因没有 3′－OH 的存在，所以无法将缺少的一段补起来(图 11－44a)，这样一来，新合成的链会逐步减短。但实际上 T7 噬菌体双链的 DNA 长度是恒定的，否则便不能存活，那么 T7 采用了什么策略解决这一问题呢？1972 年 J. D. Watson 提出了串联体(concatemer)假说，解决这一难题。由于在 T7 侵染的细菌中确实分离到了 T7 的串联体，因此这一模型也得到了承认。串联体模型如图 11－44b 所示：两个 5′ 有空缺的分子通过 3′ 端凸出的重复序列，

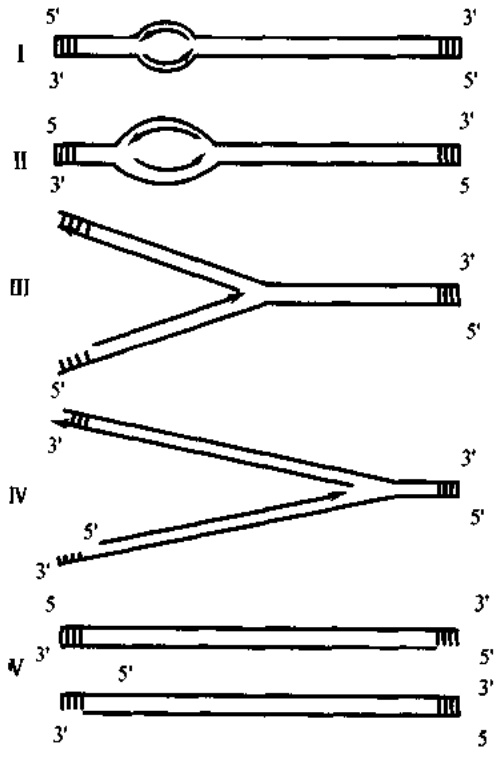


图 11－44(a)　线状 DNA 复制时 5′端可能形成空缺。

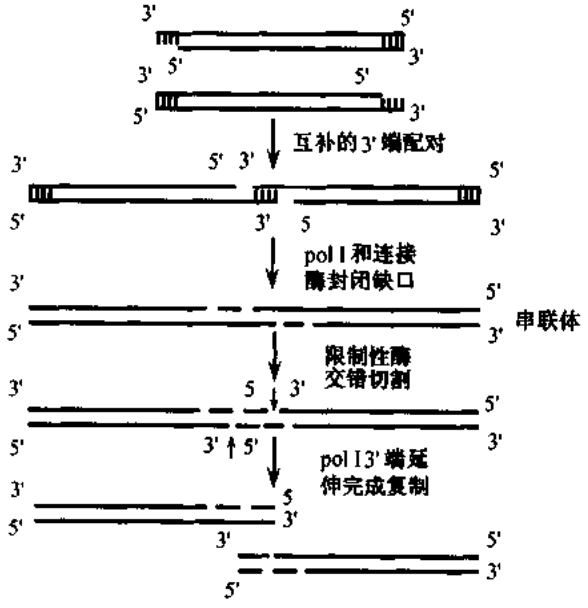


图 11－44(b)　T7 线状 DNA 复制终止的模型。

互补配对。然后 DNA 聚合酶Ⅰ从 3′ 端延伸补起这个空缺，留下最后的裂缝由连接酶封闭产生了一个大分子串联体。再由限制酶在连接处交叉切割，产生 3′ 凹端，由 DNA 多聚酶Ⅰ在 3′－OH 上延伸填满新的空缺，产生两条完整的双链。

第六节　复制的过程

一、半不连续复制

前面已阐明任何一种 DNA 聚合酶合成方向都是从 5′向 3′方向延伸，而 DNA 模板链是反向平行的双链，这样在一条链上，DNA 合成方向和复制移动方向相同（前导链），而在另一条模板上却是相反的（后滞链）。那么在复制叉中新链是如何合成的呢？1968 年冈崎（Okazaki）及其同事进行了一系列实验，回答了这一问题。

他们做了两组实验（图 11－45），一组是**脉冲标记实验**（pulse－labeling experiment）。以 *E.coli* 为材料，在培养时，培养基中加入同位素[^{3}H]标记的 dTTP，经 30 秒后，DNA 刚开始复制，分离 DNA，然后在碱中沉淀，变性，让新合成的单链和模板链分开，再用 CsCl 密度梯度离心，以沉降的快慢来确定片段的大小，再检测放射标记，结果表明被 ^{3}H 标记的片段，也就是新合成的片段，沉降系数为 20S 左右，即都是长 1000～2000nt 的 DNA 片段，而亲本链要比它大 20～50 倍；第二组实验是脉冲追逐实验。目的是检测早期合成的 DNA 片段以后的命运又是如何的？是依然如旧的短片段，还是连接成了长片段。这个实验是先进行标记培养 30 秒，以后除去同位素继续培养几分钟，再分离 DNA 在碱中沉降，检测结果。先合成的（带标记）的 DNA 不再是短片段，而和总 DNA 的情况相似，沉降系数为 70S～120S 较长的片段，这意味着刚开始合成的片段都是短片段，以后再连接成长片段，人们就把最初合成的短片段称为冈崎片段（Okazaki fragment）。

由于这个实验的结果使得人们普遍认为：无论是前导链还是后滞链都是先合成小片段，然后在连成大片段，称为“不连续复制”（discontinuous replication）。然而由模型预测应该只有一半放射标记存在于小片段中，但实验的结果却全部是小片段 DNA，为什么从前导链模板的 3′－OH 端延伸会不连续合成呢？人们在思考这个问题。1978 年 Olivera 提出了半不连续（semidiscontinuous）复制模型，也就是说前导链上的合成是连续的，只有后滞链上的合成才是不连续的。现在已经弄清原来是由于细胞内都存在有 dTTP 和 dUTP，而 DNApol Ⅲ却并不能区分它们，因此也会将 dUTP 加入到 DNA 中，形成 A·U 对。那么在 DNA 中为什么没有 U 的存在呢？这是因为 *E.coli* 细胞里有双重“保险”，防止了 U 的“混入”。第一道关是细胞里的 dUTPase，它能使 dUTP 变成 dUMP，dUMP 是不能作为 DNA 合成的底物，这样它就不再能加入 DNA 中。但总还有些漏网之“鱼”，它们还是逃过了此关，混入 DNA 中，这就靠第二道关来清除“异己”，这道关的主角是尿嘧啶 N－糖苷酶（uracil N-glycosylase），它可以切断混合尿苷的糖苷键，形成无 Pu 和无 Py 位点（apurinic or apyrimidinic，AP），再由 AP 内切酶在 AP 位点切出一个缺口，进一步进行切除修复（见第 21 章）（图 11－46）。在细胞内尿嘧啶 N－糖苷酶作用较快，而 AP 酶作用较慢，在新链合成之初约 1200bp 就有可能掺进一个 U，但很快就被尿嘧啶 N－糖苷酶切断糖苷键，在 AP 酶未作用前在脉冲标记实验中就提取了 DNA，用 NaOH 沉淀时，AP 位点易断裂，所以前导链也成了小片段。以下实验证实了这个解释：

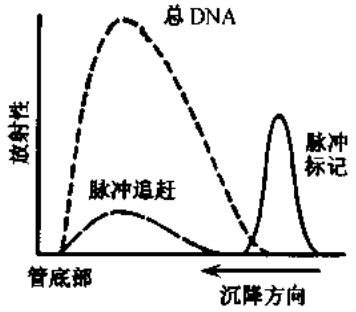


图 11－45　脉冲标记和脉冲追踪。(仿 D. Freifelder: Molecular Biology, 1983, Fig. 8－26a)

(1)在 *dut*⁻突变体(dUTPase 缺失)中冈崎片段比在 *dut*⁺中为短。这是因为 U 掺入机会增加;

(2)在 *ung*⁻(尿嘧啶 N－糖苷酶缺失)突变体中，新合成的 DNA 约有一半由片段组成。这是因为尿嘧啶 N－糖苷酶缺失，不会切除 U 的糖苷键，也就不会出现 AP 位点，所以碱沉淀时不易断裂，从而保持了半不连续的原貌。

(3)在 *dut*⁻，*ung*⁻双突变体中，结果和实验(2)相同，更进一步证实了此推测。

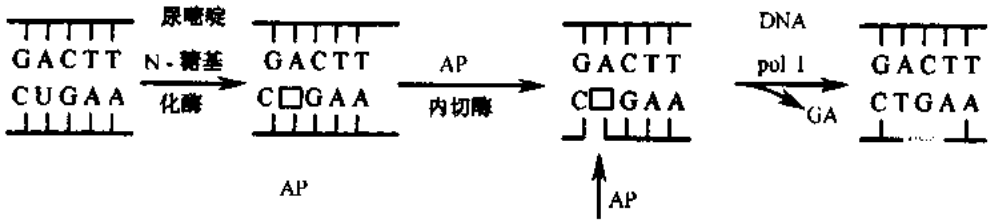


图 11－46　在 DNA 合成中 U 的切除和修复。

二、DNA 聚合反应的特征

(1) DNA 的聚合反应是以 dNTP 为底物，以带有引物的 DNA 为模板，按碱基互补配对的原则在 3′－OH 上加一个 dNTP;

(2) 需二价正离子如 Mg^{2+}，由引物提供的 3′－OH 作亲核进攻形成磷酸二酯键(图 11－47,48);

(3) 链沿 5′→3′方向延长;

(4) 碱基互补配对，若错配则被先切除掉;

(5) 反应具有进行性(processivity)，DNA pol 从 DNA 上解离下来前可催化多次反应。

三、链的延伸

在 *E. coli* 的复制叉中，由拓扑异构酶解螺旋，并使解旋产生的正超恢复成负超，由解旋酶切开氢键，打开双链(图 11－49)，然后 SSB 蛋白结合在单链上，使其稳定，暂不复性。复制叉上两条单链分别为模板，以 5′→3′方向合成新链，链的合成方向和复制叉移动方向相同的是前导链(leading strand)，相反的是后滞链(图 11－49,50)。前导链由 DNA pol Ⅲ在最初的 RNA 引物的 3′－OH 一直连续延伸，直到完成。而在后滞链上合成甚为复杂。由 PriA,B,C, DnaB,C, DnaT 及引发酶结合形成引发体，合成 2－5nt 的引物 RNA，DNA pol Ⅲ在引物 RNA－3′－OH 上加入 dNTP 合成 1～2kb 的冈崎片段，然后由 DNA pol Ⅰ用 5′→3′外切酶活性切除引物，再从冈崎片段的 3′端合成短片段 DNA，填补空缺，相邻两片段之间的缺刻由连接酶封闭，完成半不连续复制(图 11－51)。

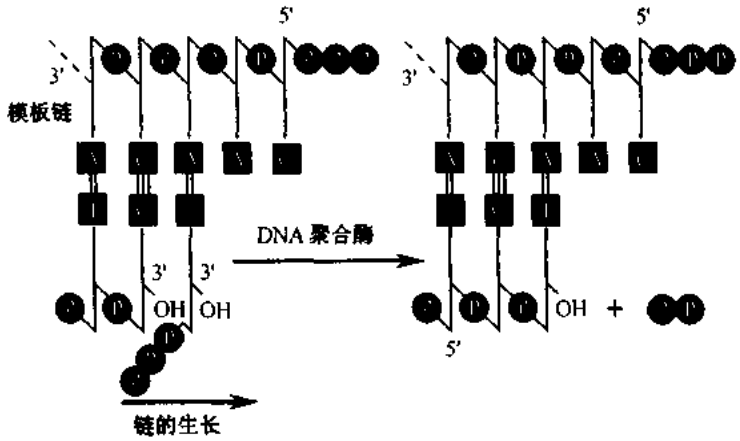


图 11－47 DNA 的聚合反应。(转引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 11－3b)

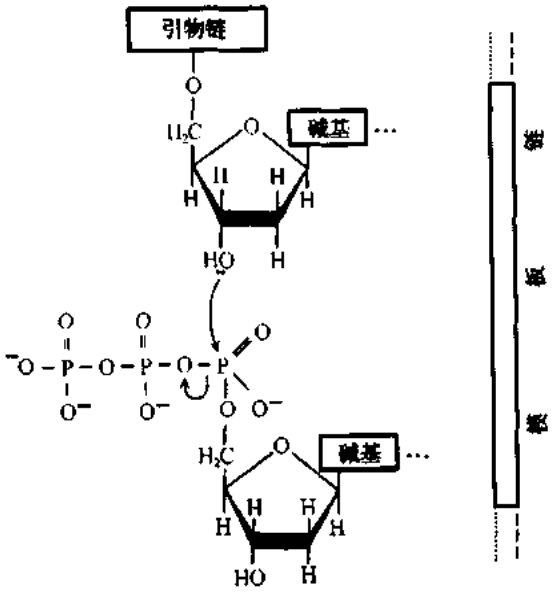


图 11－48 DNA 聚合酶催化链的延伸。(转引自 Griffiths A.J.F. *et al*: An Intiroduction to Genetic Analysis, 5th ed., 1993, Fig. 11－26)

四、复制体和回环模型

在复制中按理前导链上 DNA pol Ⅲ 的移动方向和复制叉移动方向相同，而在后滞链上其移动方向相反，也形成了一个矛盾。后来人们发现：①每一个复制叉含有特别大的多聚体Ⅲ复合体(见图 11－52)，它是一个不对称的二聚体，可能同时催化两条链。②在电镜下观察到双螺旋 DNA 聚合酶 α 也同样具有两个活性部位，推测双螺旋复制也是同时复制的。

所谓**复制体**（replisome）是由解旋酶、引发酶和 DNA pol Ⅲ 全酶组成的复合体，DNA 合成时，沿着复制叉的方向移动。后滞链模板形成了一个回环，使环中 RNA 引物和冈崎片段的合成方向与前导链一致，以适应双链在同一复制体上进行复制，此模型称回环模型(图 11－52)。这一模型有以下特点

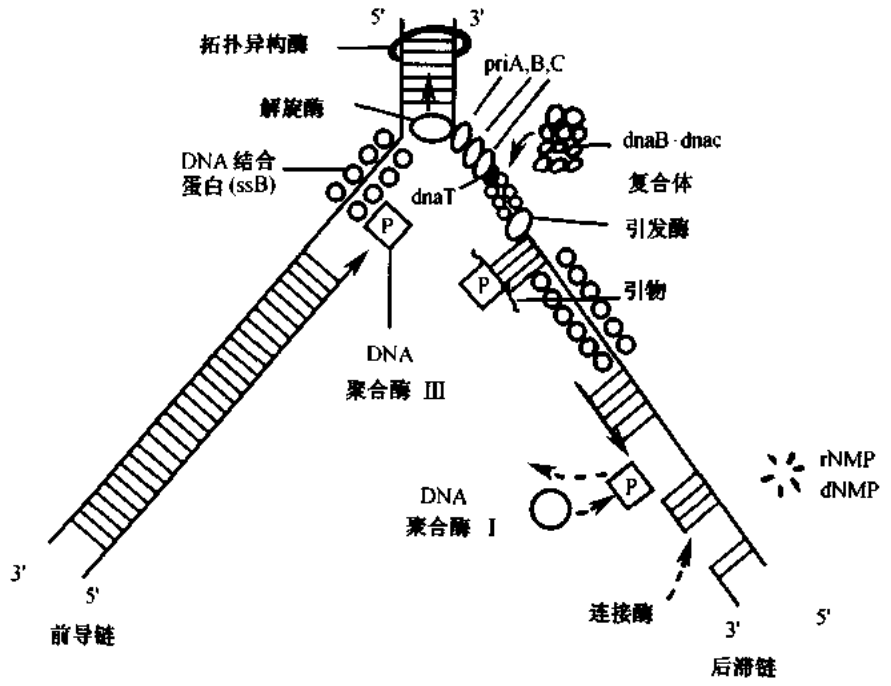


图 11 - 49　DNA 复制叉。(A. Kornberg and T. A. Baker: DNA Replication, 1992. 2nd ed.)

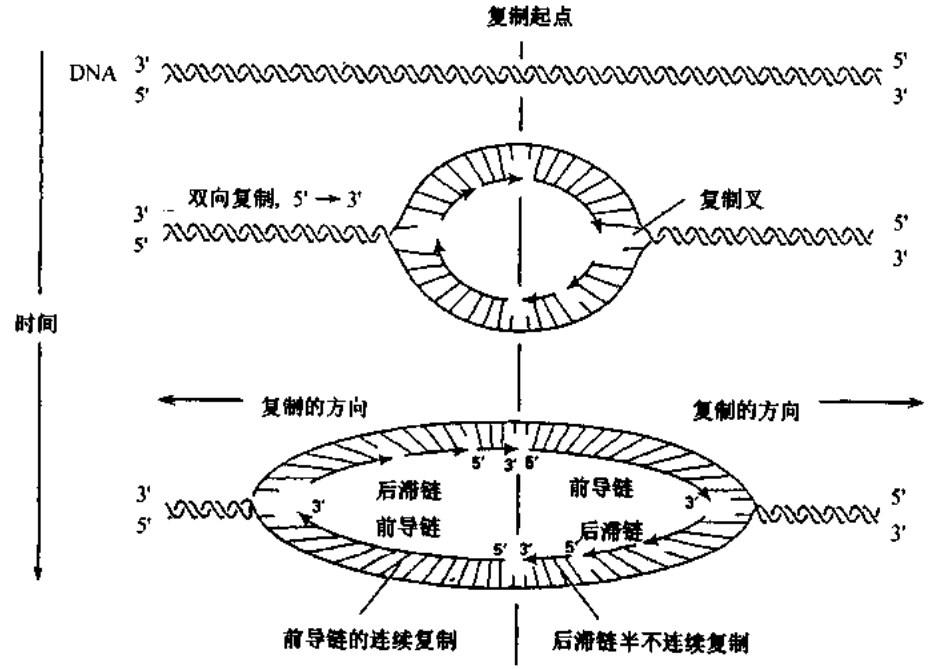


图 11 - 50　DNA 的双向复制。(仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 11 - 6)

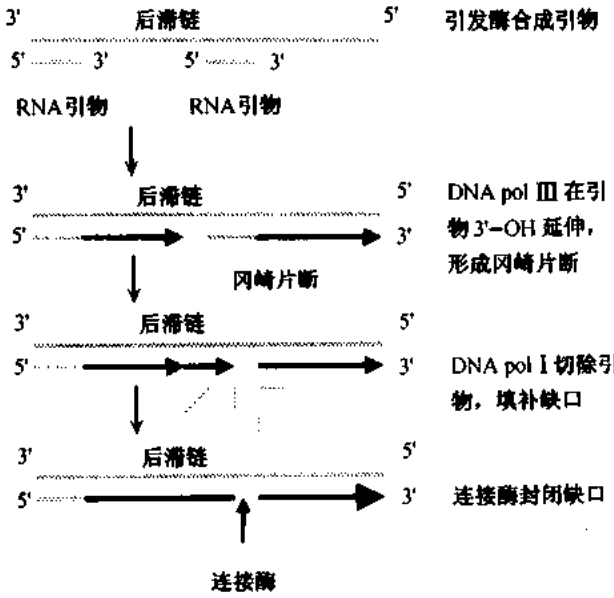


图 11－51　后滞链的合成。

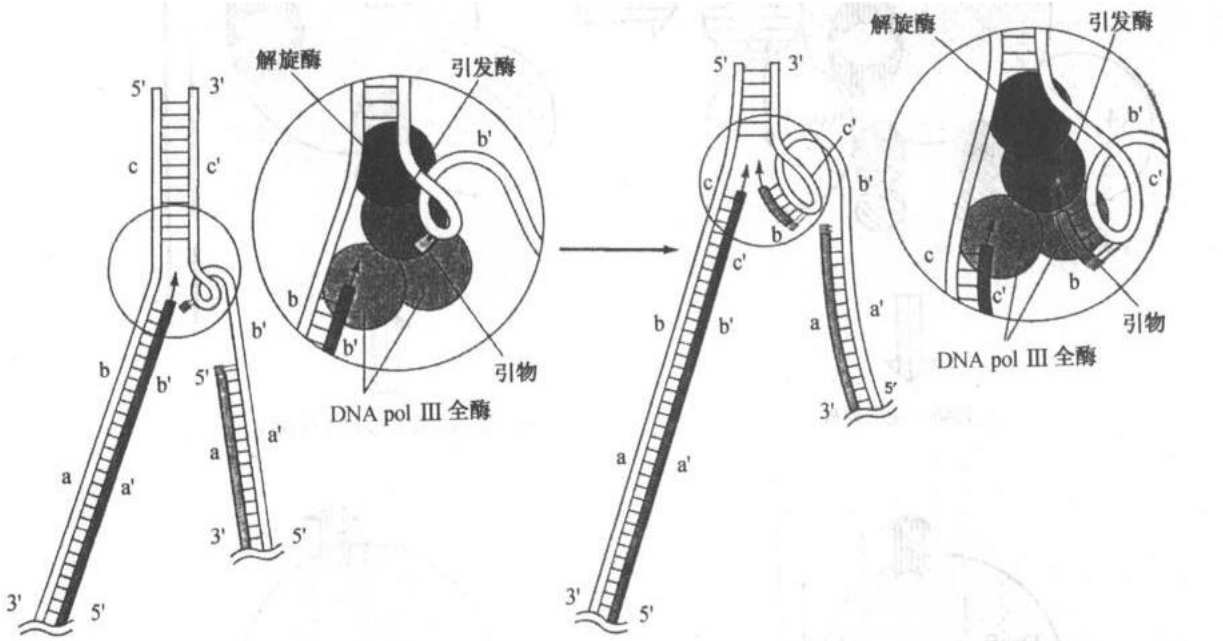


图 11－52　后滞链的模板成环使作为二聚体的 DNA pol Ⅲ在复制叉上能同时合成两条子链。(转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 11－34)

(1) 聚合酶和引发体联系在一起可以协调它们的作用；

(2) 复制体中的 DNA pol Ⅲ是二聚体；

(3) 后滞链是形成回环复制的。

根据 DNA pol Ⅲ 的不对称结构，人们进一步推测复制叉的结构。两个催化核心分别和 DNA 的两条模板链结合，全酶沿着前导链模板随着复制叉的移动而移动，而后滞链模板从复制体上“拉出”一段，形成一个环。DnaB 产生了解链点，并沿着后滞链以 5′→3′方向移动。DnaB 与 DNA pol 的一个亚基连接(可能后滞链上聚合酶的 τ 亚基)。这个连接的作用增加了 DNA 的合成速度。在缺乏这种连接时，DNA pol 跟着解旋酶以每秒约 35nt 的速率前进。若 τ 与 DnaB 连接速率将增加 10 倍。前导链的合成为后滞链

的合成提供了一个单链的环。当解旋点推进时这个环变得更大。冈崎片段开始合成后，后滞链上的酶核心复合体将单链模板拉向β亚基形成的"夹子"。在后滞链上聚合酶在合成一个冈崎片段之前，单链至少需要延伸一个冈崎片段那么长。

当合成冈崎片段时为什么会产生环呢？当 pol Ⅲ全酶在 DNA 上遇到一个裂口时核心复合体和滑动的"夹子"离开，然后在别的地方再和新的β亚基结合(图 11－53)，表示在冈崎片段末端发生这种反应。当核心酶合成完一段冈崎片段时，核心复合体就会和模板解离，将环释放出来，另一个(也可能是同一个)核心复合体再和一个β夹子

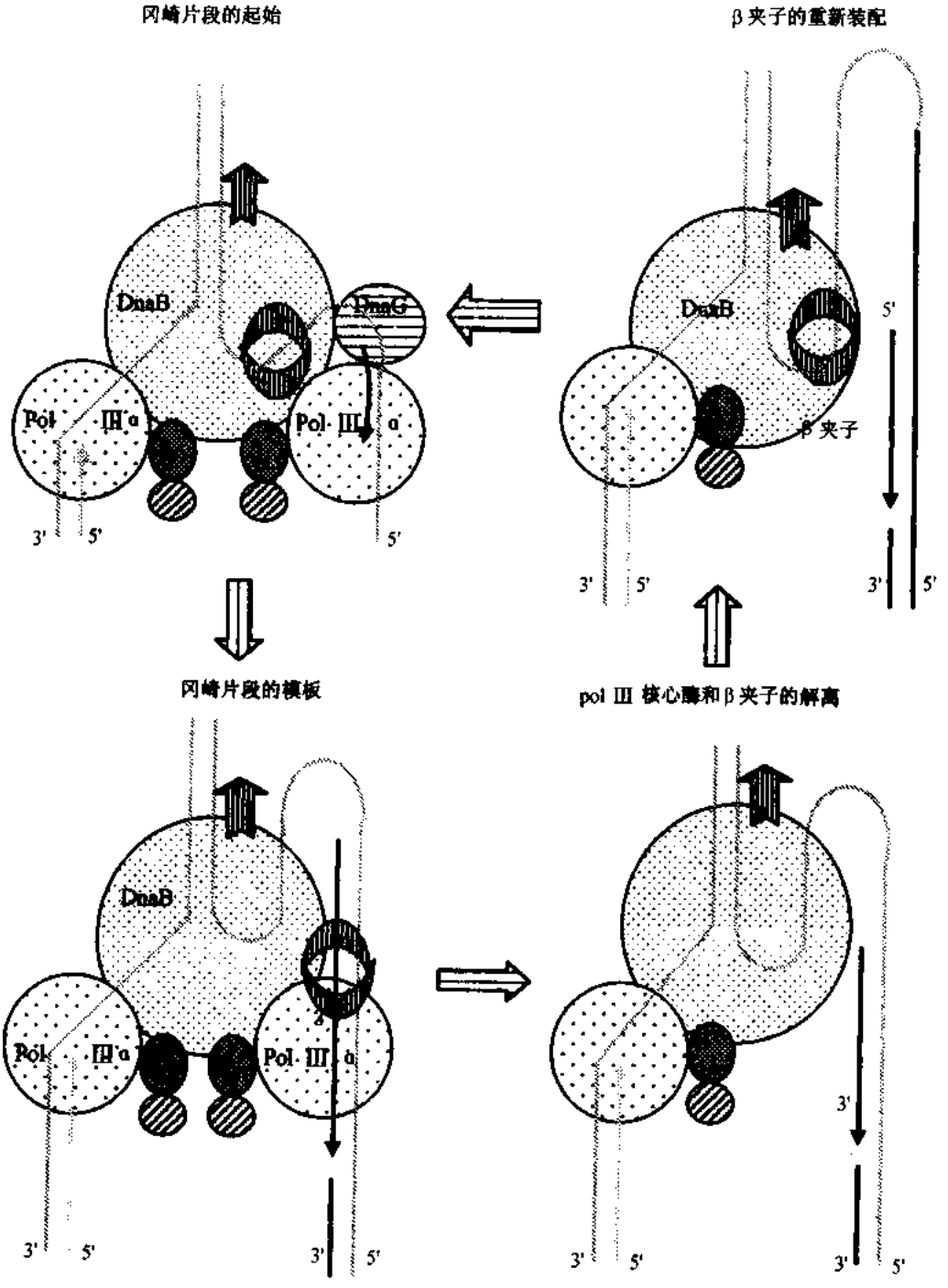


图 11－53　pol Ⅲ核心酶和β夹子在冈崎片段合成后解离，然后再重新装配。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig. 15－16)

结合，起始下一个冈崎片段的合成。有可能一个新的β"夹子"已存在于起始位点。而失去了核心复合体的β"夹子"将从模板上解离下来，然后再被重新使用。

第七节 真核生物的 DNA 复制

一、真核生物 DNA 聚合酶及有关蛋白

真核 DNA 复制有关的酶和蛋白如表 11－9 所示。真核的聚合酶有 α、β、γ、δ 和 ε 五种(表 11－10)，根据以下证据认为 α 是复制的主要酶类。

表 11－9 真核生物五种 DNA 聚合酶

DNA 聚合酶	α(Ⅰ)	δ(Ⅲ)	ε(Ⅱ)	β	γ
位置	核	核	核	核	线粒体
功能	后滞链的合成和引发	前导链的合成	修复	修复	复制
相对活性	80%			10%～15%	2%～15%
分子量	300k	170～230k	250k	40k	180～300k
亚基	催化核心(180k)两个引物酶(60,50k)一个未知	催化核心(125k)一个未知(25k)需 PCNA(37k)	催化核心一个未知	催化	催化
3'→5'外切	－	＋	＋	－	＋
双脱氧 T	不影响	不影响	弱	抑制	抑制
蚜黄素	抑制	抑制	抑制	不影响	不影响

摘自 B. Lewin: Gene V, 1994, Table 19.3

表 11－10 真核复制需要的酶和蛋白

蛋白	亚基	功能
DNA 聚合酶 α/引发酶	4	后滞链合成
DNA 聚合酶 δ	2	前导链合成
PCNA	1	延伸
复制因子 C	3	延伸(ATP)酶
复制因子 A	3	单链结合
拓扑异构酶Ⅰ和Ⅱ		维持 DNA 的拓扑结构

摘自 B. Lewin: Gene V, 1994, Table 19.4

(1)所有 α 酶的抑制剂如蚜黄素(aphidicolin)都可以抑制真核细胞的 DNA 复制，而 β、γ 的抑制剂对复制的影响不大；

(2)α 酶的温度敏感突变株的 DNA 复制也是温度敏感；

(3)用微注射的方法将 α 酶的单克隆抗体注入细胞，则可抑制细胞 DNA 的复制；

(4)α 酶的活性在细胞周期中会不断变化，而在 S 期其活性最高，此期正是 DNA 合成的时期。

DNA pol α 是作为一种复合体存在，它含有 180kDa 的催化亚基和另一小亚基相连，小亚基含有两个小蛋白，具有引发酶活性，以上两种活性是功能偶联的。若解偶联(产生 60k 和 50k 小亚基)就不能产生有酶催化活性的引物酶。此复合体在后滞链上合

成 DNA(图 11 - 54)。

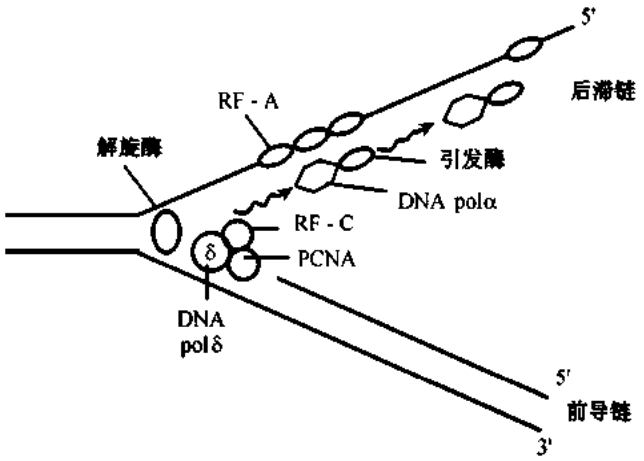


图 11 - 54　真核 DNA 的复制模型。

DNA pol δ 可能是合成前导链的。δ 酶由两个亚基组成，分子量分别为 125kDa 和 25kDa。具 3′→5′外切酶活性的聚合酶，它还需要一种附属蛋白，分子量 37kDa，这种附属蛋白可以增加 DNA pol δ 与模板/引物的相互作用，促进合成 poly dA/oligo dT 的活性。现已知道这种蛋白是分裂细胞核抗原(proliferating cell nucleus antigen PCNA)。下例的实验证实附属蛋白就是 PCNA。

(1) 两者的氨基酸顺序相同；

(2)水解的双向电泳图谱相同；

(3)都可促进 δ 酶的活性；

(4)都可在体外促进 SV40 的复制。

另一实验证明它是和复制有关的蛋白：

(1) 用荧光标记 PCNA 的抗体，分析 3T3 培养细胞，G_1/S 两期过渡时核中可以看到荧光，S 期最强，到 G_2 期荧光又消失了。

(2) 用双向电泳比较分裂细胞和静止细胞的蛋白质成分，结果发现仅分裂细胞中才有 PCNA，现已知 PCNA 是周期蛋白 (Cyclin) 中的一种，其反义 RNA 可以抑制培养细胞。

(3) 在 SV40 体外复制系统中 PCNA 是必需的一种蛋白。

(4) PCNA 能促进 pol δ，但对 pol α 无作用，可见其是 δ 酶的附属蛋白。

DNApol α 与 δ 有很多不同之处，归纳起来有以下几点：

(1) pol δ 有 3′→5′外切酶活性；

(2) 两者对不同模板的合成效率不同，α 酶在合成 poly d(A - T)时活性差。

(3) 分子量不同，α 的催化亚基为 165kDa，δ 的催化亚基为 125kDa；

(4) 纯化两种酶在不同的洗脱峰中；

(5) PCNA 对 α 酶无促进作用；

(6) 合成 poly dA/oligo dT 时 PCNA 的存在，pol α 进行性短，pol δ 进行性大。

对于真核的复制，1987 年 Miami 大学的 Downey 提出了一个模型(图 11 - 55)。他的依据是：

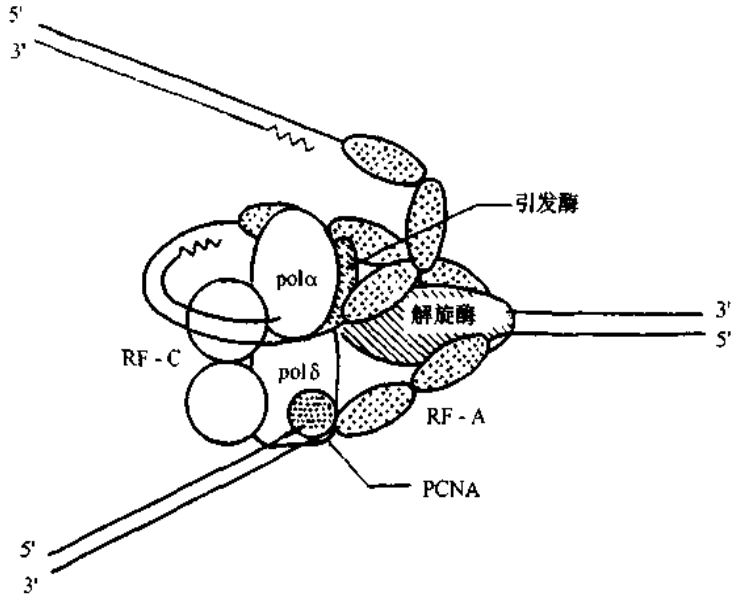


图 11 - 55 真核 DNA 复制的模型。

(1) pol δ 进行性比 pol α 大得多,所以 pol δ 合成前导链,pol α 合成后滞链。

(2) 在 PCNA 存在是 pol δ 能进行链的替代,有利于打开复制叉,而 α 酶无替代能力。

(3) pol δ 不具有引发酶的活性, pol α 常和引发酶结合在一起, 这对后滞链中形成多个 RNA 引物是适合的。

二、核小体的全保留复制

真核的染色体和原核不同, 还要涉及核小体, DNA 进行了半保留复制, 那么它又怎样和组蛋白结合形成子染色体,现在要面临以下几个问题。

在不久之前人们都认为核小体是全保留复制的,实际上却并非如此。在复制时亲代 DNA 链必定要分离, 至少 30nm 的纤丝的结构必然要破坏掉, 甚至 10nm 的纤丝也不能保全。人们都很想知道染色质的结构破坏到什么程度。这种染色质结构有什么可观察到的差别呢? 复制的过程是短暂的,这给分析复制区的结构带来了很大的困难。复制叉的结构是很特殊的。它对微球菌核酸酶具有抗性, 可被降解成大小不同的带, 形成核小体 DNA。此提示在 DNA 复制中有大的蛋白复合体与之结合。复制后核小体或多或少要立即重建。

染色质的复制并不因为组蛋白的游离而拖延合成的时间。一旦 DNA 被复制核小体便立即在两条子链上迅速形成。我们通过电镜照片(图 11 - 56)就可以了解这一点。

组蛋白怎样和 DNA 结合立即产生核小体这是一个多年来令人迷惑的问题。组蛋白是预先形成的一个八聚体围绕在 DNA 的周围, 还是一个 $H3_2 \cdot H4_2$ 四聚体作为一个核与 DNA 先结合,然后再加入 H2A · H2B 二聚体呢? 受到装配颗粒沉淀的限制,在体外的自我装配是一个很慢的过程。要想知道可模仿的生理条件是很困难的。在体外装配核小体有两条途径,如图 11 - 57 所示。

附属蛋白涉及到帮助组蛋白和 DNA 结合。在非洲爪蟾的卵中发现了这种蛋白的候选者。爪蟾卵的抽提物能介导组蛋白和外源 DNA 装配成核小体。此卵内含有两种蛋

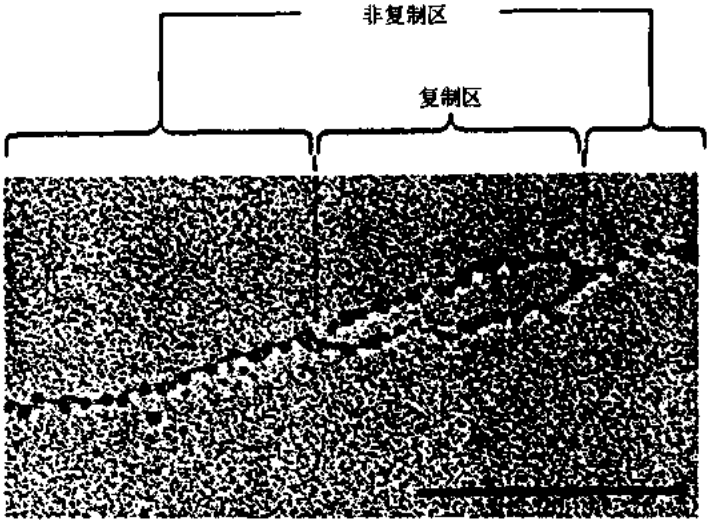


图 11－56　在染色质的复制区上两条子链上都含有核小体。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig. 27－25)

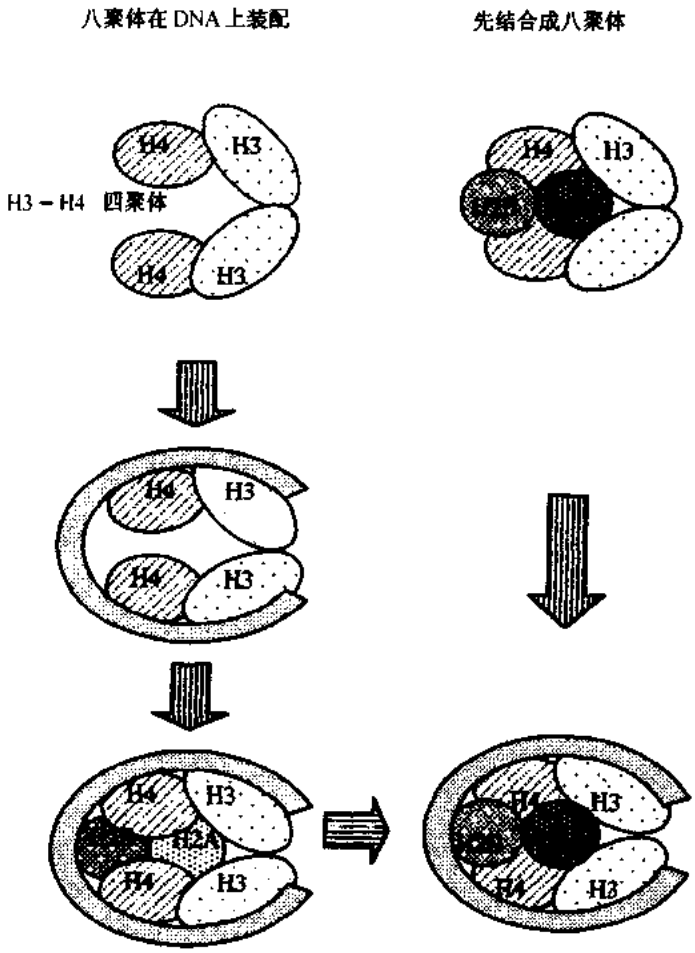


图 11－57　在体外 DNA 既可以直接和完整的组蛋白八聚体相互作用，也可先和 H3－H4 两个四聚体装配，然后再和 H2A－H2B 装配。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 27－26)

白可结合组蛋白。核质蛋白(nucleoplasmin)结合 H2A 和 H2B，N1 蛋白结合 H3 和 H4。这两种蛋白的抗体可以抑制抽提液在体外组装核小体。这些附属蛋白的功能是什么呢？他们可能起到一种“分子伴侣”的作用。它们以一种可调控的方式将单个的组蛋白或它

们的复合体[(H3·H4)或(H2A·H2B)]与 DNA 结合。这可能是必然的,因为组蛋白是碱性蛋白,它们对 DNA 有很高的亲和性。而核质蛋白和 N1 是酸性蛋白,它们和组蛋白结合减少了负电荷。这种相互作用使组蛋白可以形成核小体而无需成为另一种动力学中间体。

在体外产生核小体的尝试开始是将游离 DNA 和组蛋白进行组装，但核小体在体内形成时仅仅在 DNA 复制的时候。人们用人类细胞的抽提物建立了一种模仿系统，使 SV40 DNA 复制并装配产生染色质。这个装配反应优先发生在复制的 DNA 上。这需要一种辅助因子，CAF－1，它含有五个以上的亚基，总分子量 238kDa。CAF－1 的作用是化学剂量性的。其功能是和新合成的 H3，H4 结合，它储存在复制叉上，形成的核小体很快地延伸到 200bp，但它们不含 H1 组蛋白，表明没有 H1 的存在也能留下适当的间隔。

当染色质复制时,一段经复制的 DNA 已和核小体结合了,产生两条子链。在复制区核小体是怎样预先存在的呢?是组蛋白八聚体解离成游离的组蛋白还是保持装配的状态呢？完整的八聚体通过组蛋白的交联是可以被检测出的。图 11－58 表明一个实

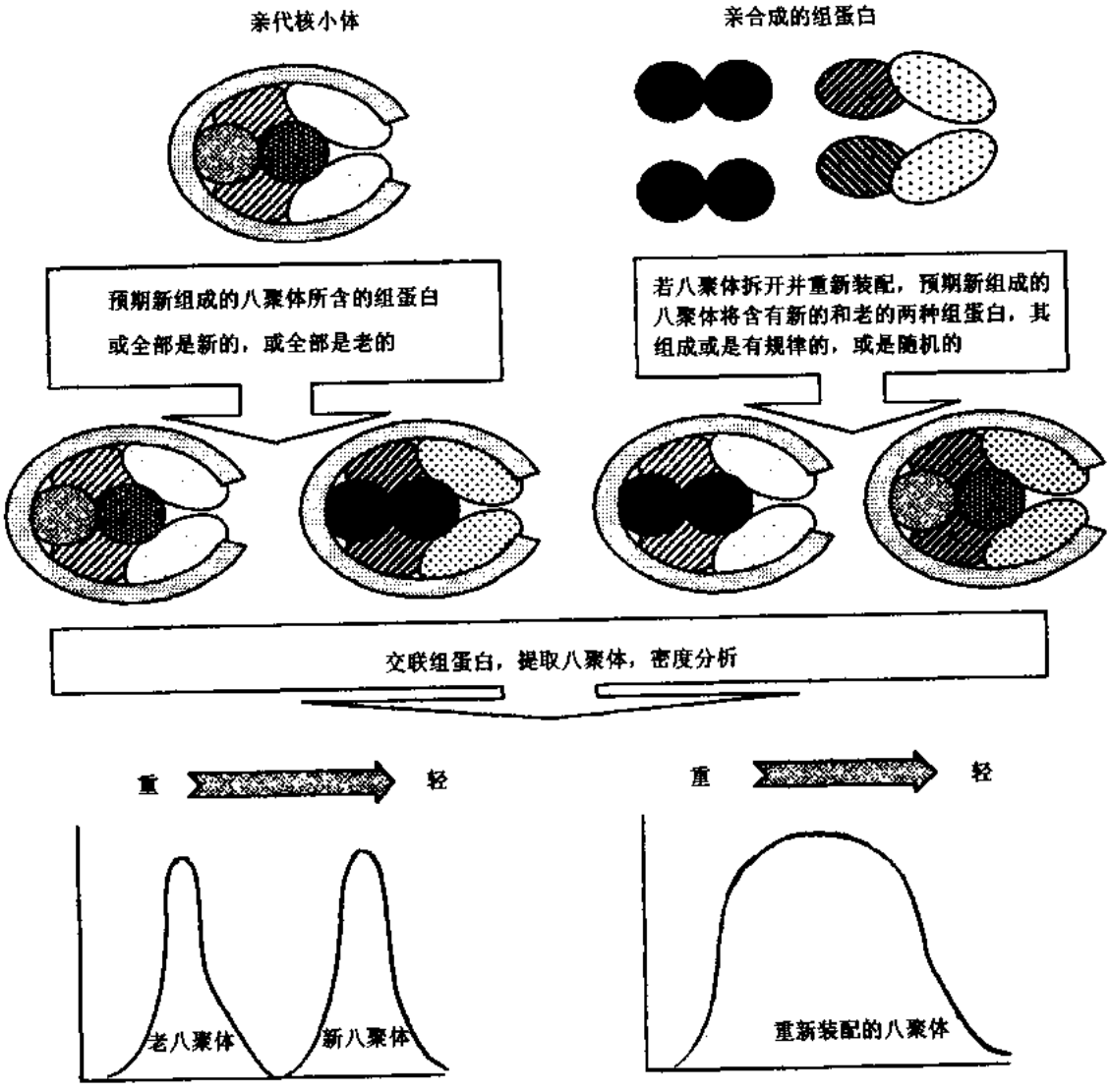


图 11－58　组蛋白的同位素标记实验，表明组蛋白八聚体的复制不是全保留的。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 27－27)

验，即在复制前将细胞放在含有重标记氨基酸中培养生长，以此鉴别原来的组蛋白。复制时放在轻标记氨基酸的条件下进行。此时组蛋白被交联并经密度梯度离心。若老的八聚体是全保留的话，那么此时它们应在高密度位置，而新合成的八聚体应当在低密度位置，但如果老的八聚体是半保留的话将呈中等密度的。实验结果在高密度仅有少量的物质，表明组蛋白八聚体的复制不是全保留的。

核小体八聚体解聚和重新装配的模式还不十分清楚，有可能完全解离成组蛋白组分。一种可能是八聚体丢失了一个或两个 H2A · H2B 二聚体，它们被新合成的或老的所取代，在这种情况下 $H3_2 \cdot H4_2$ 四聚体是保守的。有可能在转录时相似的情况就发生了，此 $H3_2 \cdot H4_2$ 四聚体在复制时可能具有与单链结合的能力，这可能增加了它保留在前导链上的可能性。这就意味着至少有的组蛋白是在 DNA 上组装的。

三、SV40 复制的起始

SV40 病毒的 DNA 为双链环状，有核小体结构，全长 5243bp。具复制起始区（图 11－59）最低限度是从第 5208 到第 29 个碱基共 64bp 的区域，保留这段序列，SV40 就具有 40%～50%的复制能力，此区主要含三个不同区域：10bp 的前期区，27bp 的 T 抗原结合位点和 17bp 的 A－T 丰富区。在 T 抗原结合位点Ⅱ有回文序列存在。完整的复制起始区是从第 5229bp 开始到第 30 个碱基，具有 100%的复制活性。与前者相比仅增加 T 抗原结合位点的大部分区域，此区域含有两个顺向重复序列。

起始时在复制原点上发生一系列的反应和 *E.coli* 的 oriC 及 λ 的 ori λ 相似。由 SV40 自己编码的两个六聚体 T 抗原结合到 DNA 起始位点的回文序列上，在 ATP 存在时改变 DNA 结构，使双链打开，此解链区很短，并不是 A · T 丰富区，但它有一个特点，一条链上几乎全是嘌呤，另一条链几乎全是嘧啶。A · T 区与此区相邻，当 T 蛋白和 DNA 结合时 A · T 区会弯曲。T 抗原本身具有解旋酶活性，所以并不需要相当于原核中 DnaB 蛋白类似物，T 抗原由 708aa 组成，结构如图 11－60 所示，其识别的序列为 GAGGC 重复序列。SV40 的复制起始区还需要其他的蛋白因子。

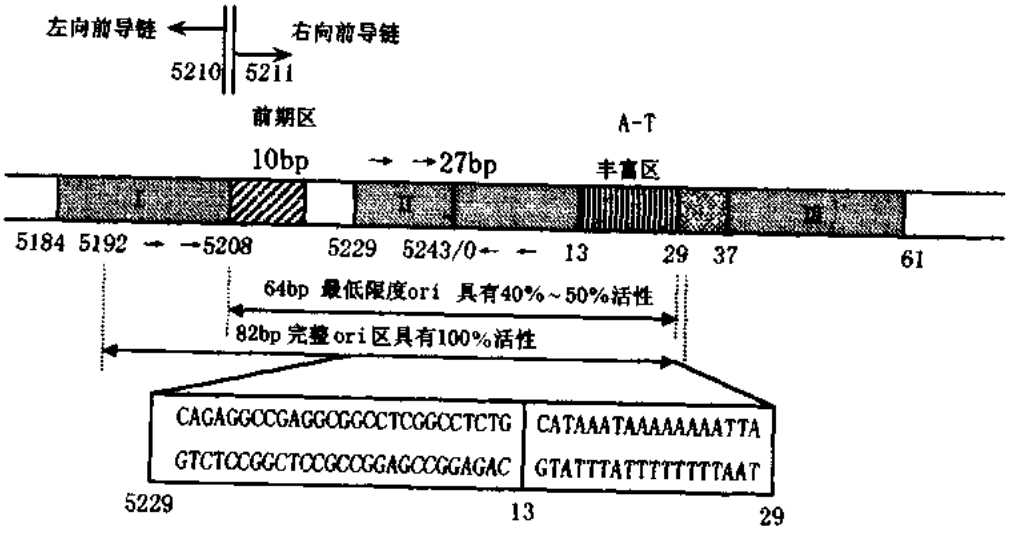


图 11－59　SV40 复制起始区的结构。

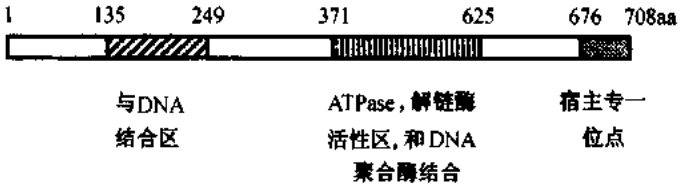


图 11－60　T抗原的结构。

RF－A　是从人培养细胞 293 中分离出的一种 SV40 体外复制所需要的因子。它可以和 T 抗原，Topo Ⅰ、Ⅱ及 ATP 形成复制起始复合体。它本身是一个三个亚基的复合物(为 70kDa，34kDa 和 11kDa)。其功能是真核生物的单链结合蛋白，比结合双链的亲和力高 100 倍。

RF－C　也是从 SV40 体外复制模型中鉴定出的另一种蛋白质因子，位于核内。由四个亚基组成(两个 41kDa，两个 37kDa)，具有 ATP 酶的活性。DNA pol δ 与 RF－C 作用结合到前导链的第一个引物的 3′－OH 上，继续延伸。它和 polα 亚基结合在后滞链上合成冈崎片段。PCNA 和 RF－C 有协同效应，它可激活 RF－C 的活性(引物结合蛋白)，因此在后滞链上。

起始的过程可以分以下几步(图 11－61)：

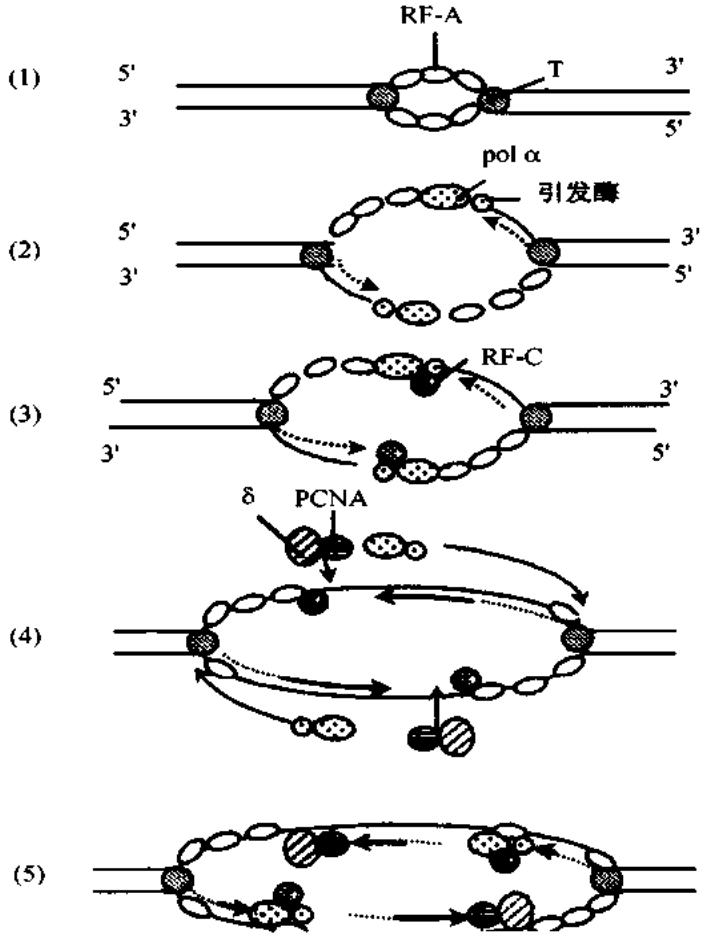


图 11－61　SV40 DNA 复制的起始。

(1) T 蛋白具有解链酶活性，通过水解 ATP 解开双链，RF－A 立即和单链结合，使之稳定；

(2) 复制叉移动一段距离，polα－引物酶复合物结合到复制区，合成第一段 RNA 引物；

(3) RF－C 结合到引物上，使 polα 合成第一条冈崎片段；

(4) 到前导链和后滞链的分界区，polα 从单链上解下来，再结合到新的复制叉处合成后滞链；

(5) polδ，RF－C 和 PCNA 结合到前导链的第一条冈崎片段 3′端开始合成前导链；

(6) 根据电镜观察的结果，SV40 的复制可能也形成复制体，后滞链形成回环。

四、真核 DNA 复制的特点

每一条真核生物的染色体都是一条线状双链 DNA 分子，若每条染色体只有一个复制起点的话，那么要完成整套染色体的复制要花费很多时间。例如人类单倍体基因组（23 条染色体）有 2.75×10^9bp，平均每条染色体长约有 10^8bp，在人的细胞中复制的速率为 2kb/min，那么将要花费 830 小时才能完成一条染色体的复制，实际上人的细胞周期为 72 小时。若每个细胞的周期要比人类胚胎发育时间还要长的话，那么妊娠的时间就需要许多年而不是“十月”怀胎了。

真核 DNA 复制的真实速度要比以一个 ori 推算的结果快得多。例如果蝇胚胎二倍体整个染色体只需三分钟就可以复制完，这比 *E. coli* 的染色体复制快六倍。即使果蝇的 DNA 长度要比 *E. coli* 的大 100 倍。果蝇 DNA 复制之所以这样快，主要是因为在其基因组中有多个复制起点，在真核生物中一般都是如此。

在原核生物中第一轮复制尚未结束前第二轮复制又从 ori 区开始了，此称为成熟前起始（prematureinitiation），而在真核生物一般情况下要等第一轮复制结束后第二轮才开始。

真核的复制起点到两边的复制终点称为一个**复制子**(replicon)或一个复制单位(replication unite)。

表 11－11 中表明细菌和真核生物的复制子数目、平均长度和复制叉移动的速率。值得注意的是真核和原核复制子的大小是非常相似的，但复制的速率真核要比原核慢得多。这是因为在一般情况下每个复制的真核染色体都有很多复制子。真核的复制起始点的数目和复制单位的大小是随着细胞的特点变化而改变的，如不同的生长速率的细胞和不同组织的细胞。图 11－62 表明果蝇 DNA 片段上有大量的复制子。图 a 表示电镜照片，DNA 长 500kb 左右(约为 *E. coli* DNA 的 1/10)且有多个复制子存在。

表 11－11　细菌和真核生物复制的比较

生物	复制子数	复制子长度	复制速率
细菌(*E. coli*)	1	4200kb	50kb/min
酵母(*S. cerevisiae*)	500	40kb	3.6 kb/min
果蝇(*D. melanogaster*)	3500	40kb	2.6 kb/min
爪蟾(*X. laevis*)	15000	200kb	500 bp/min
小鼠(*M. musculus*)	25000	150kb	2.2 kb/min
蚕豆(*V. faba*)	35000	300kb	

P. J. Russell: Genetics, 3rd ed. 1992, Table 11.3

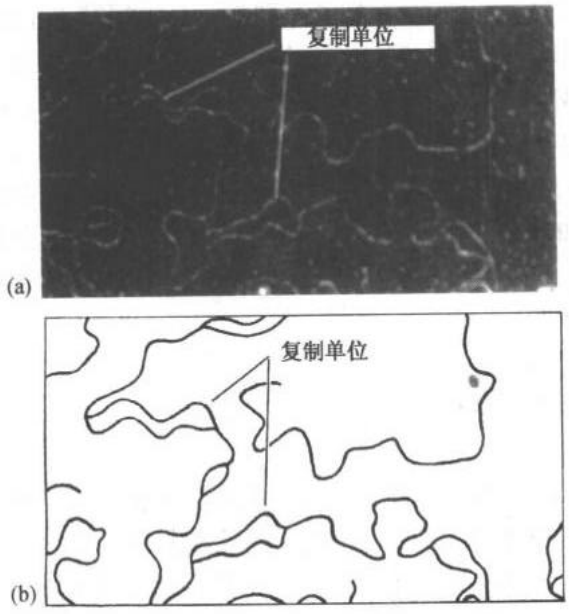


图 11－62　果蝇 DNA 的复制。(a)复制单位的电镜照片(复制子)(b)根据电镜照片绘制的图。(转引自 Russell, P.J.:Genetics,3rd ed.,1992,Fig. 11－21)

在生物的基因组中,并非所有的复制子都同时进行 DNA 复制。而是有一定的起始时间顺序。不同类型的细胞复制起始的顺序也不同,现在尚不清楚控制复制早迟的分子机制。图 11－63 是放射自显影研究的结果,表明一段 DNA 上的三个复制子总是在不同的时间开始复制。当两个相邻的复制叉延伸,融合后染色体就变成了两条染色单体。在一般情况下随着一组复制起点的同步激活,染色体 DNA 片段开始复制。

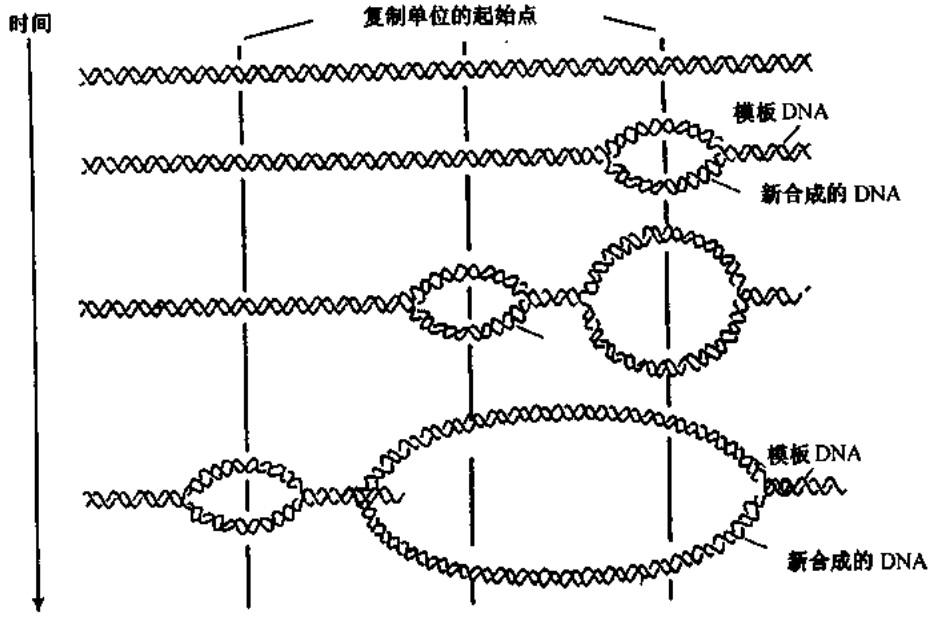


图 11－63　真核生物染色体上复制单位的复制起始顺序。(转引自 Russell,P.J.:Genetics,3rd ed.,1992,Fig. 11－22)

在真核生物中复制起始区有没有特殊序列呢？在啤酒酵母中已鉴别出特殊序列，将它作为一个部件来构建一个环状的 DNA 分子，这个核外 DNA 分子在酵母中是可以自主复制的。这个序列称为自主复制序列(autonomously replicating sequences)或叫 ARS 元件，在其他的真核生物中也发现类似于酵母 ARS 元件的序列。

人们常常误以为 ARS 元件在细胞中仅和复制有关，实际上在细胞中某些元件的复制起始作用是明确的；但从另一面来说也并非是在染色体片段上的所有 ARS 元件都用于复制起始，比较一下酵母中多种 ARS 元件可发现有 11bp 的保守序列：

5′ATTTATPuTTTA 3′

3′TAAAYAPyAAAT 5′

这个 A－T 丰富区顺序是和复制起点的功能相适应的，因为这段序列的双链易于变性。

在不同的发育时期，真核的复制起始点和复制子的大小是会改变的，有些复制起点在有的发育阶段就不再发挥作用。如果蝇受精后复制原点从五千个上升到五万个。在发育早期到成体时每个复制子长仅 7.9kb，也就是说复制子较多，而发育到成体时每个复制子长 40kb，毫无疑问，其中很多 ori 区不再起作用。

真核 DNA 复制时 DNA 聚合酶有五种(α、β、γ、δ 和 ε)，低等真核生物有三种(α、ε 和 δ)，其中只有 δ 酶具有 3′→5′外切酶活性；而原核生物只有三种 DNA 聚合酶 (pol Ⅰ、Ⅱ、Ⅲ)，且都具有 3′→5′外切酶活性。真核的 pol δ 和 α 分别负责合成前导链和后滞链，而原核的 pol Ⅲ是两者兼顾。

在真核细胞内 DNA 引物酶和 pol α 紧密偶联，而在原核细胞内引发酶是和解旋酶偶联在一起，形成复制体的一部分。

五、真核复制的终止

真核生物的染色体也是线形的，复制终止时也像 T7 那样会面临 5′空缺的问题，在真核染色体中又是如何解决这个问题的呢？

前面(第 10 章)已经介绍了真核染色体末端有一个特殊的结构——端粒。是一个短片段的重复结构，而且末端还形成特殊发夹，它是如何复制的呢？这还得从末端的重复来源谈起。

原生动物有两个核，大核和小核。大核为营养核，小核为生殖核。1983 年人们发现纤毛原虫在有性生殖过程中大核由小核发育而来，其中小核 rRNA 基因为单拷贝，它被切下来成为游离基因，经扩增达到一万个拷贝，有趣的是这些 rDNA 小分子末端都有端粒，即(CCCCTT)n 重复片段，但小核 rDNA 基因的两侧原来是没有这段重复顺序的，显然这些末端重复是在切下以后加上去的，如何来证实这个推论呢？人们做了以下各种实验逐步搞清了重复末端的来源，发现了一种特殊的复制酶，叫做端粒酶(telomerase)。

1984 年 J. Shampay 等构建了线性质粒，两端连接了四膜虫的端粒，转化到酵母后，这一线性质粒的两端仍加上了酵母的端粒，当将线性的质粒一端连接四膜虫的端粒，另一端连接酵母的端粒后，转化酵母发现有四膜虫端粒的一端仍加上酵母的端粒，而在有酵母端粒的另一端保持不变，这就是有名的加尾实验(图 11－64)，结果表明端粒复制的模板并不在端粒上。那么模板在何处呢？

1985 年 Blackburn 进行了体外实验来寻找端粒复制酶的线索。他用四膜虫无细胞

抽提物在体外进行了端粒的“加尾”实验，即以四膜虫末端重复顺序（G_4T_2）4 为引物，以同位素标记的 dGTP 和 dTTP 为底物进行“加尾”，结果抽提物可以使引物延伸几百个 nt，且以六个 nt 顺序（G_4T_2）为单位往返重复，既然末端序列不作为模板，但延伸的序列又和它相同，那么加尾和末端结构到底有什么内在的联系呢？这种复制酶是否需要识别呢？

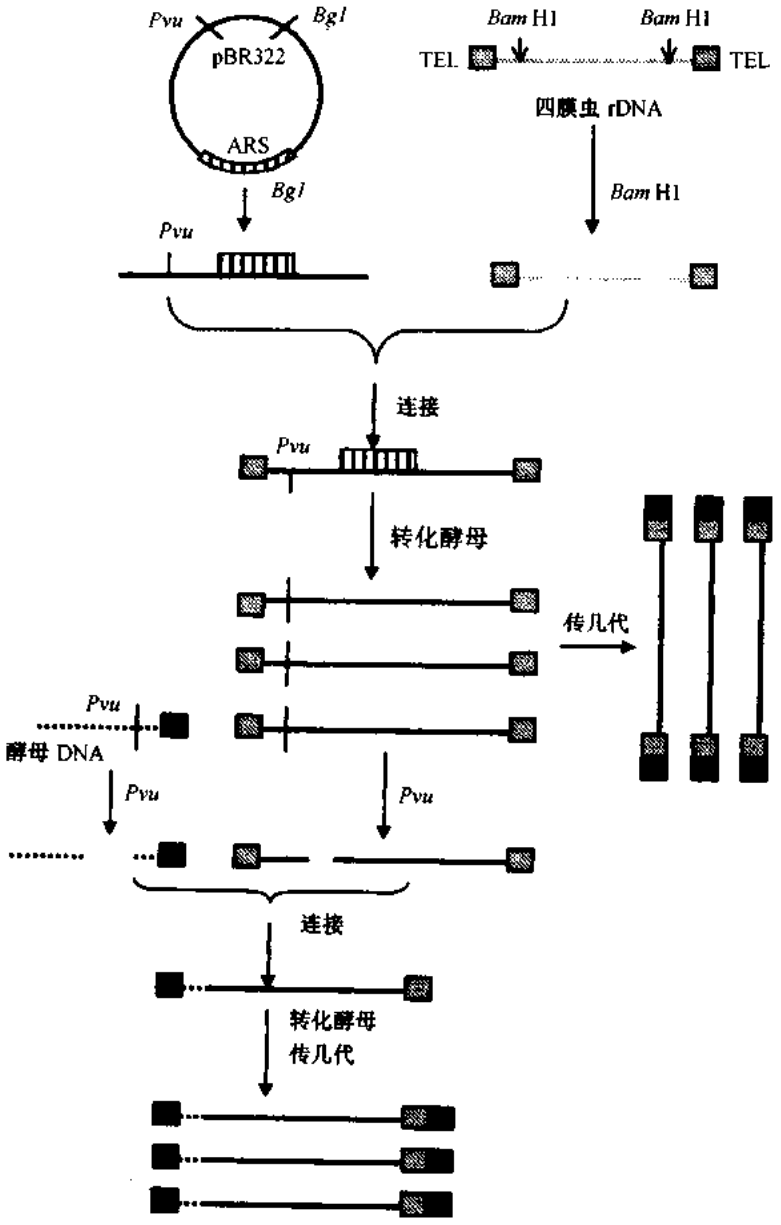


图 11－64　四膜虫与酵母 TEL 的拼接以及酵母细胞将其特有的 TEL 加到四膜虫的 TEL 上。（参考 Shampg, J and Blackburn *et al.*: Nature, 1984, 310: 154～157）

人们又改变条件继续实验，当以酵母的末端重复顺序为引物时，用四膜虫抽提物加上的仍是四膜虫的端粒；当以 pBR322 的片段为引物时，则不能延长末端，当以四膜虫的另一条互补链（C_4A_2）4 为引物时也不能延长；当把反应基质加热到 90℃或用蛋白酶处理也不再能延长末端。这表明①端粒的加尾方向是按富含 G 链 5′→3′延伸，每次

以(T_2G_4)为单位加上去的;②端粒本身不作为模板,而是作为引物从头合成;③酶系能识别拟被"加尾"的末端顺序;④酶是蛋白质性质的。人们称其为端粒转移酶,现称端粒酶。既然末端的顺序和"加尾"有关,那么其特异性是什么呢?

1990年 Yao M C 等成功地在四膜虫体内诱导了外源 DNA 序列的加尾,证明特殊染色体的断裂序列有引导端粒酶结尾作用。同年 Yu(余国良)和 Blackburn 报导四膜虫端粒酶可在 A－T 丰富区的非端粒顺序后面加上端粒。这就解释了在大核发育中,染色体断裂后可以直接加尾的现象。

从这两个实验看来端粒酶在体内对引物的要求并不严格,不是一定要有富含 G 的重复顺序为引物。但尚未表明末端和酶的直接关系是什么?即酶的作用机制是怎样的?

1987年 Greider 和 Blackburn 就分离纯化了端粒酶,其分子量在 200～500kDa。进一步分析弄清了端粒酶是一种含 RNA 的蛋白质,RNA 长 159nt,其中第 46、47、48 位 nt 被

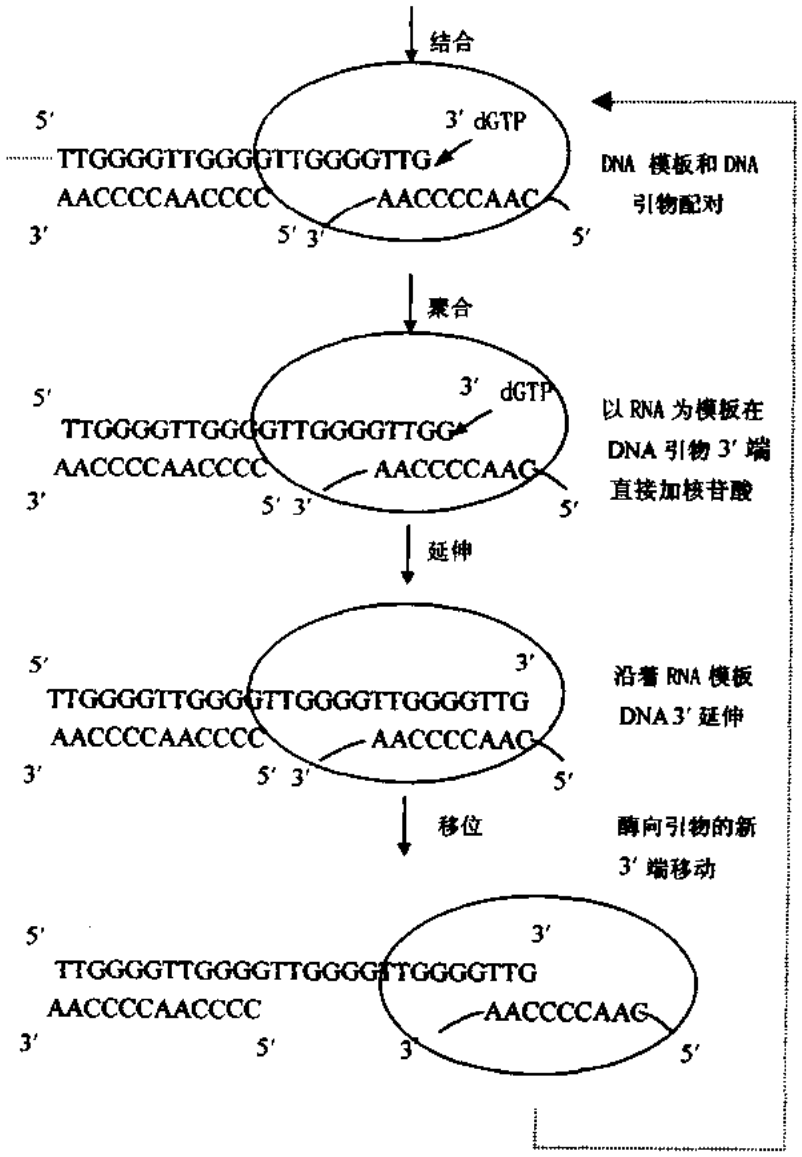


图 11－65　端粒酶复制机制模型。(仿 Greider and Blackburn: *Nature*, 1989, 337: 331－337)

修饰过，核酸酶对这三个位点不能剪切，而这一位点恰是 CAACCCCAA 的序列区。

这 159nt 的 RNA 中有没有功能区呢？人们采用了点突变、缺失和反义技术来加以研究。首先是用 159nt 的不同片段的反义寡核苷酸来处理端粒酶，结果发现反义序列 5′GCACTGATTTTGGGGTTG3′可强烈地抑制此酶的加尾活性。此抑制作用存在着与 $d(T_2G_4)_4$ 的竞争性。即将反义序列与 $d(T_2G_4)_4$ 片段同时加入端粒酶反应溶液中，反义序列只能起部分的抑制作用。说明反义序列（其 3′端有一个 3′GTTGGGGTT5′结构）能妨碍 $d(T_2G_4)_4$ 与端粒酶的接近并加尾，因此提出端粒酶 RNA 中的 5′CAACCCCAA3′序列是此酶的活性部位，并为 (T_2G_4) 的末端重复提供模板。

1990 年余国良等对四膜虫端粒酶中的 RNA 基因的核心序列 5′CAACCCCAA3′进行了定点突变，再转化四膜虫，结果受体细胞后代的染色体两端出现相应突变的端粒序列，从而有力地证实了端粒酶 RNA 的核心序列是复制端粒 DNA 的模板。

Greider 和 Blackburn （1989 年）提出了端粒的复制模型（图 11－65）。反应为四步进行，①首先是端粒酶与端粒 DNA 结合，端粒酶中的 RNA 与凸出的 3′引物配对；②以 RNA 为模板，在 DNA 3′端上从头合成 DNA；③延伸六个 nt；④合成一个重复单位后，端粒酶向 DNA 模板新合成的 3′移动，引物再和模板配对，就这样循环往复，周而复始。最后延伸的 3′端再回折，以 G－G 配对的方式形成发夹结构（图 11－66），产生端粒。

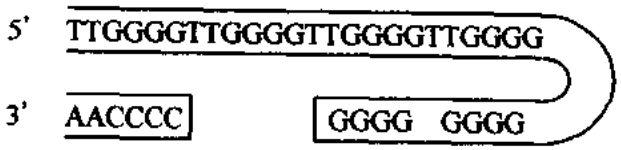


图 11－66 端粒 3′延伸，末端回折形成发夹结构。

六、复制的调控

目前对复制的调控了解得很少，远不如转录和翻译调控研究得那么深入具体。

细菌在不同的培养条件下，细胞培养的时间不同，毫无疑问和复制有关，也就是说在不同的条件下 DNA 复制受到了调控。看来调节的关键在于原核 DNA 复制起始蛋白和酶的合成速率以及复制起点的结构状态，这也涉及到转录激活的情况。从带有 OriC 的质粒在体外复制的情况看，其复制起始的频率主要取决于模板的转录活性和含有 Dna 蛋白、Hu 蛋白等的复制起始复合物形成。这些蛋白的浓度直接和复制有关。

真核的复制是和细胞周期紧密相联的，只有在 S 期 DNA 才进行复制，因此细胞周期的调节也是对 DNA 合成的调节。在真核生物中这方面研究得比较多的是酵母。在 G_1 期酵母的 *CLN1*、*CLN2* 和 *CLN3* 基因（在人类中 *Cyclin C*, *D*, *E* 基因）合成了 G_1－周期蛋白（cyclin）作为调节亚基和 p34 蛋白结合形成促成熟因子（maturation-promoting factor, MPF），并激活了 p34 蛋白的蛋白激酶活性，从而激活了一些转录因子和 Oct1，这些转录因子识别 cdc6（编码 DNA 聚合酶），cdc8（胸苷激酶），cdc9（DNA 连接酶），cdc21（胸苷酸合成酶）等 5′端的保守顺序 5′ACGCGTNA3′并促使它们转录；Oct1 可激活 H2B 的启动子，

为 DNA 复制做好物质准备，细胞周期越过 S 期的限制点，DNA 开始合成。在合成中各种酶和蛋白的浓度都会影响合成的速率。

目前对复制调控研究得比较清楚的是质粒 ColE1(图 11－67)，它的 DNA 复制是受到 Rop(或 Rom)蛋白(63aa)和反义 RNA 的调控。

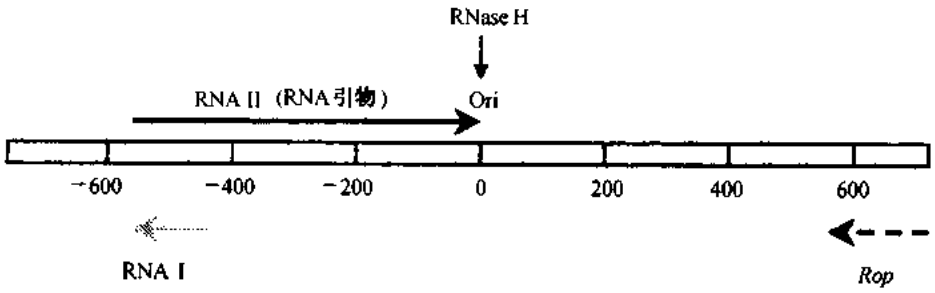


图 11－67　ColE1 质粒的复制调控。

ColE1 复制起始时先从 ori 上游的－555bp 处开始右向转录，产生 RNA Ⅱ，当它向右延伸时在 ori 区被 RNase H 切断，产生 3′－OH，作为 DNA 复制的引物。但在上游－445 处还有另一个左向转录起始点，产生了一个小分子的 RNA，叫做 RNA Ⅰ，它和 RNA Ⅱ的 5′端互补，所以也是一个反义 RNA。当 RNA Ⅰ和 RNA Ⅱ分子 5′端互补结合时产生了二级结构，形成了三个互补的茎环。这时 RNase H 就不能识别和切割 RNA Ⅱ也就不能在 ori 区产生 3′－OH 提供引物，因此抑制了引物的形成。在 ori 下游有一个 Rop 基因，它编码的蛋白可以促进 RNA Ⅰ和 RNA Ⅱ互相结合，并使 RNA Ⅱ合成到 100～220bp 即终止，这样达不到 ori 区，也无法提供引物。在这个调控系统中 RNA Ⅰ和 Rop 蛋白都是负调控因子。RNA Ⅱ是正调控因子。

第十二章 转录

基本概念

转录是在原核和真核细胞中以 DNA 为模板合成 RNA 的过程。

在原核和真核生物中，转录过程是相似的，包括 DNA 变性，RNA 聚合酶结合在单链 DNA 上以 5′→3′方向合成 RNA 分子等过程。双链中只有一条链作为转录模板，合成单链 RNA 分子。启动子和终止子序列决定转录的起始和终止。

在 *E.coli* 中一种 RNA 聚合酶转录各种 RNA(mRNA，tRNA 和 rRNA)。在真核细胞中有三类不同的 RNA 聚合酶，它们的功能不同。RNA pol Ⅰ转录 4 种 rRNA 中的 3 种；RNA pol Ⅱ转录 mRNA 和一些 snRNA；RNA Ⅲ转录第四种 rRNA，tRNA 以及其余的 snRNA。

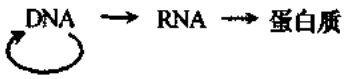
3 种真核生物的 RNA 聚合酶，与 *E.coli* RNA 聚合酶不同，不直接和启动子区结合，而是通过转录起始因子的介导来起始 RNA 的合成。对于每一种 RNA 聚合酶来说，转录因子是特异的，它可以识别启动子的特殊序列。

蛋白质编码基因的启动子位于转录起始位点的上游，由不同组合的启动原件所构成。特异的转录因子和调节因子结合在这些原件上，促进 RNA pol Ⅱ转录起始。增强子离启动子较远，它可被调节因子识别结合，具有促进基因转录的功能。

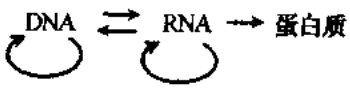
由 RNA pol Ⅲ转录的基因的启动子，位于下游，在其基因编码序列内部。这种启动子，根据所转录的 RNA 的种类，由不同的功能区组合而构成。转录因子识别这些功能区，促进 RNA 聚合酶转录起始。

18S、5.8S 和 28S rRNA 作为一个转录单位—道转录，产生前体 RNA 分子。大部分真核生物的 18S、5.8S 和 28S rRNA 都是以串联重复排列，每个重复单位被不转录的间隔序列（nontranscribed specer，NTs）所分隔。转录单位的启动子位于 NTS 中，其功能是和特异的转录因子相结合，促进 RNA pol Ⅰ的转录起始。

从孟德尔定律问世以后，人们就知道了生物的各种性状是由基因控制的。一基因法测指出了遗传信息的传递方向；一酶学说的建立进一步地明确了基因是以酶的形式通过控制生化反应链来控制的。酶或蛋白和基因又是什么样的关系呢？也就是说遗传信息怎样传递，怎样表达成性状呢？就在 Watson 和 Crick 建立 DNA 双螺旋模型后的第三年，1957 年 Crick 提出了中心法测(central dogma)，指出了遗传信息的传递方向：



1970 年 H.Temin 和 D.Baltimore 发现了反转录酶后，Crick 对中心法测又作了部分修改：



也就是说由 DNA 通过转录将遗传信息传递给 RNA，RNA 通过翻译把信息传递给蛋白(图 12－1)。通过这种单向的传递，遗传信息通过蛋白质的不同形式，如酶、结构蛋白、运载蛋白、调节蛋白等表达成一种性状。

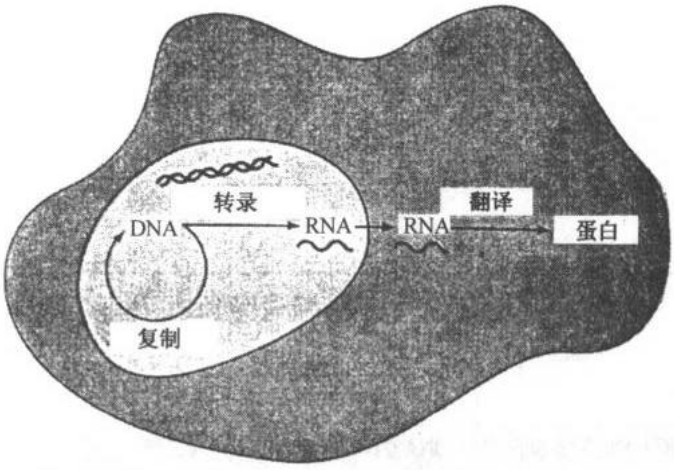


图 12－1　遗传信息传递的方向和步骤。
（引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 13－2）

第一节　信使的发现

储存在 DNA 分子中的遗传信息能在复制中产生更多的拷贝，并翻译成蛋白质。DNA 的功能构成了信息的流动，遗传信息如何转变成蛋白质呢？转录就是其中的重要的一环。基因表达时以 DNA 的一条链为模板合成 RNA，这一过程就是转录（transcription）。催化合成 RNA 的酶叫做 RNA 聚合酶（RNA polymerase）。RNA 和 DNA 结构相似，所不同之处在于：①RNA 一般以单链形式存在；②RNA 中的核糖其 C′－2 为不脱氧的；③尿苷（U）取代了 DNA 中的胸苷。细胞中的 RNA 分成三种：mRNA（信使 RNA），tRNA（转运 RNA）和 rRNA（核糖体 RNA）。它们的功能各不相同。mRNA 是合成蛋白质的模板，tRNA 是转运特异氨基酸的运载工具，rRNA 是合成蛋白质的装置。mRNA 的碱基序列，决定着蛋白质装配时氨基酸的序列。

1955 年 Brachet 用洋葱根尖和变形虫进行了实验；若加入 RNA 酶降解细胞中的 RNA，则蛋白质合成就停止，若再加入从酵母中提取的 RNA，则又可以重新合成一些蛋白质，这就表明，蛋白质的合成是依赖于 RNA。

同年 Goldstein 和 Plaut 用同位素标记变形虫（*Amoeba proteus*）RNA 前体，发现标记的 RNA 都在核内，表明 RNA 是在核内合成的。在标记追踪（pulse-chase）实验中，用短脉冲标记 RNA 前体，然后将细胞核转移到未标记的变形虫中。经过一段时间发现被标记的 RNA 分子已在细胞质中，这就表明 RNA 在核中合成，然后转移到细胞质内，而蛋白质是在细胞质中合成的，因此 RNA 就成为在 DNA 和蛋白质之间传递信息的信使的最佳候选者。

1956 年 Elliot Volkin 和 Lawrence Astrachan 作了一项很有意思的观察：当 *E. coli* 被 T2 感染，RNA 的合成迅速停止，但噬菌体的 RNA 却开始迅速合成。用同位素脉冲—追踪标记表明噬菌的 RNA 在很短的时间内就进行合成，但很快又消失了，表明 RNA 的半衰期是很短的。由于这种新合成的 RNA 的碱基比和 T2 的 DNA 碱基比相似，而和细菌的碱基比不同，所以可以确定新合成的 RNA 是 T2 的 RNA。由于 T2 感染细菌时注入的是 DNA，而在细胞里合成的是 RNA，可见 DNA 是合成 RNA 的模板。最令人信服的证

据来自 DNA－RNA 的杂交实验。Hall，BD 和 Spiegeman，S 将 T2 噬菌体感染 *E. coli* 后立即产生的 RNA 分离出来，分别与 T2 和 *E. coli* 的 DNA 进行分子杂交，结果发现这种 RNA 只能和 T2 的 DNA 杂交形成"杂种"链，而不能和 *E. coli* 的 DNA 进行杂交。表明 T2 产生的这种 RNA（即 mRNA）至少和 T2 的 DNA 中的一条链是互补的。

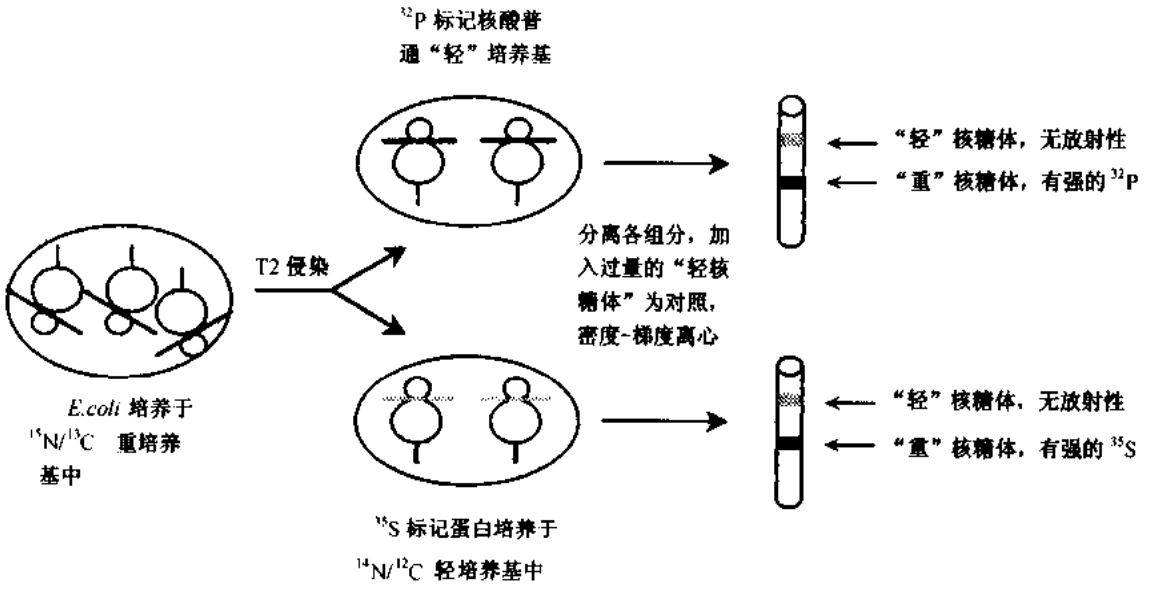


图 12－2　证明 T2 噬菌体的转录后翻译核糖体是 *E. coli* 的，无特异性。只有 mRNA 才有特异性可传递遗传信息。

Brenner，Jacob 和 Meselson（1961）进行了一系列的实验（图 12－2），他们将 *E. coli* 培养在 ^{15}N/^{13}C 的培养基中，因此合成的 RNA 和蛋白都被"重"同位素所标记。也就是说凡是"重"的核糖体，RNA 和蛋白都是细菌的，然后用 T2 感染 *E. coli*，细菌的 RNA 停止合成，而开始合成 T2 的 RNA 此时用普通的"轻"培养基（^{14}N/^{12}C），但分别以 ^{32}P 来标记新合成的 T2 RNA，以 ^{35}S 标记新合成的 T2 蛋白，因此任何重新合成的核糖体，RNA，及蛋白都是"轻"的但带但有放射性同位素。经培养一段时间后破碎细胞，加入过量的轻的核糖体作对照，进行密度梯度离心，结果"轻"的核糖体上不具有放射性，"重"的核糖体上具有 ^{32}P 和 ^{35}S，表明①T2 未合成核糖体，"轻"核糖体是后加放的。②T2 翻译时是借用了细菌原来合成的核糖体，所以核糖体并无特异性，核糖体上结合的 mRNA，其序列的特异性才是指导合成蛋白质的遗传信息，从而提出了 mRNA 作为"信使"的证据。因此他们将这种能把遗传信息从 DNA 传递到蛋白质上的物质称为"信使"。他们预言①这种"信使"应是一个多核苷酸；②其平均分子量不小于 5×10^{5}（假定密码比是 3），足以携带一个基因的遗传信息；③它们至少是暂时连在核糖体上；④其碱基组成反映了 DNA 的序列；⑤它们能高速更新。Volkin 和 Astrachan 发现高速更新的 RNA 似乎完全符合以上条件。Jacob 和 Monod 将它定名为信使 RNA（messenger RNA）或 mRNA。

第二节 RNA 的酶促合成

一、RNA 合成的基本特点

1960 年 Weiss，S，B 及其他科学家发现了一个能在 DNA 模板上合成 RNA 的酶系统，使得 mRNA 假设得到了更有力的支持。这种酶称做依赖于 DNA 的 RNA 聚合酶，简称 RNA 聚合酶(RNA pol)。几乎所有的细胞中都有这种酶，其特点是：①以核糖核苷三磷酸(rNTP)为底物；②以 DNA 为模板；③按 5′→3′方向合成；④无需引物的存在能单独起始链的合成；⑤合成时，第一个引入的 rNTP 是以三磷酸形式存在；⑥在体内 DNA 双链中仅以一条链作为模板；⑦合成的 RNA 的序列和模板是互补的。

以上这些特点都是通过一系列的实验而了解的，这些实验的设计常常是独具匠心，非常值得借鉴。

如 1963 年 J. Marmur 和 Doty 改进了 Hall 和 Spiegelnan 的杂交方法，能区分 DNA 的两条链，从而证明了在体内 DNA 只有一条链被转录。他们采用枯草杆菌(*Bacillus subtilis*)的 SP8 噬菌体为材料，SP8 的双链 DNA 很特殊，大多数的 Pu 分布于同一条链上，大部分 Py 分布于另一条链上，使两条链的“轻”，“重”差异明显。这样将其 DNA 加热到 100℃使之变性，然后用密度梯度离心就可将轻链和重链分开。在分离前，将感染 SP8

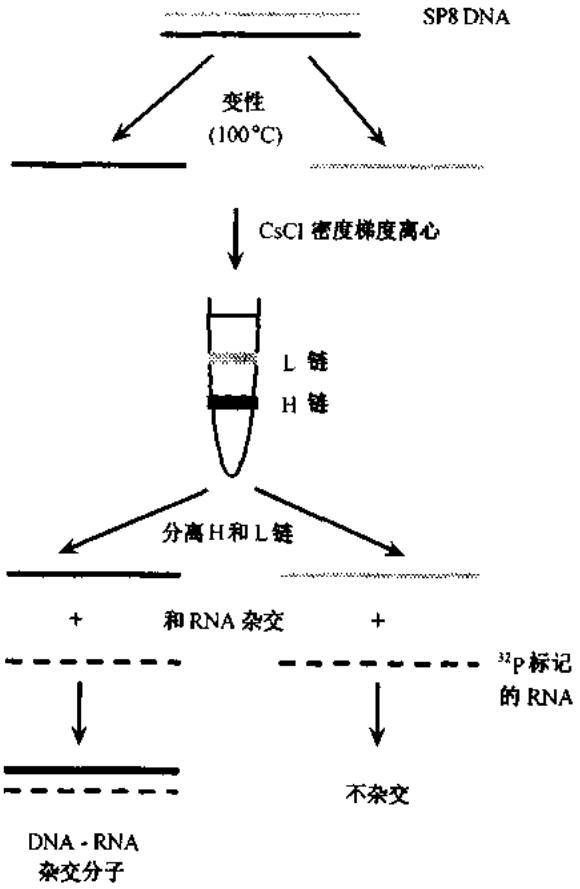


图 12 - 3　SP8　DNA 仅一条链(H)和其 mRNA 可以杂交。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 12 - 4)

噬菌体的枯草杆菌放在含有 ^{32}P 的培养基中培养，使新合成 SP8 的 RNA 标记上 ^{32}P，然后将标记上 ^{32}P 的 mRNA 分别和重链及轻链杂交，结果 mRNA 只和 DNA 重链形成 DNA－RNA“杂种”分子(图 12－3)。显然，只有 SP8 的重链主可作为其 mRNA 的模板。现在人们将这条作为转录模板的 DNA 单链称为模板链或反义链(antisen strand)，非模板链称为有义链（sense strand）或编码链。在体外 DNA 的两条链都可作为 RNA 合成的模板。

无论是在体外还是在体内，mRNA 的合成总是延着 5′→3′方向进行的。其证据是来自以下的实验：人们发现 *E.coli* 在不同的培养温度下 mRNA 合成的速度不同，在 0℃时需 13 秒钟才能加上一个核苷酸，但在 37℃就要快得多，每秒就可加上 40 个核苷酸，人们就利用这个差别以 ^{14}C 来标记 U 来培养 *E.coli*，这样就可能有区别地标记了伸长的 mRNA 的两端。当提取这种正在伸长的 mRNA 分子经分析发现 ^{14}C 标记首先出现在伸长的 3′端，因此可以证明合成是延着 5′→3′方向进行的。

RNA 的合成和 DNA 的复制其化学反应十分相似(图 12－4)，但也有一些重要的差别：①转录时只有一条 DNA 链为模板，而复制时两条链都可作为模板；②转录时 DNA－RNA 杂合双链分子是不稳定的，RNA 链很快被游离的 DNA 所取代，DNA 又恢复双链状态，RNA 合成后会释放出来，而 DNA 复制叉形成后一直打开，不断向两侧延伸，新合成的链和亲本形成子链；③RNA 合成不需引物，而 DNA 复制一定要有引物存在；④转录的底物是 rNTP，复制的底物是 dNTP；⑤聚合酶系不同。

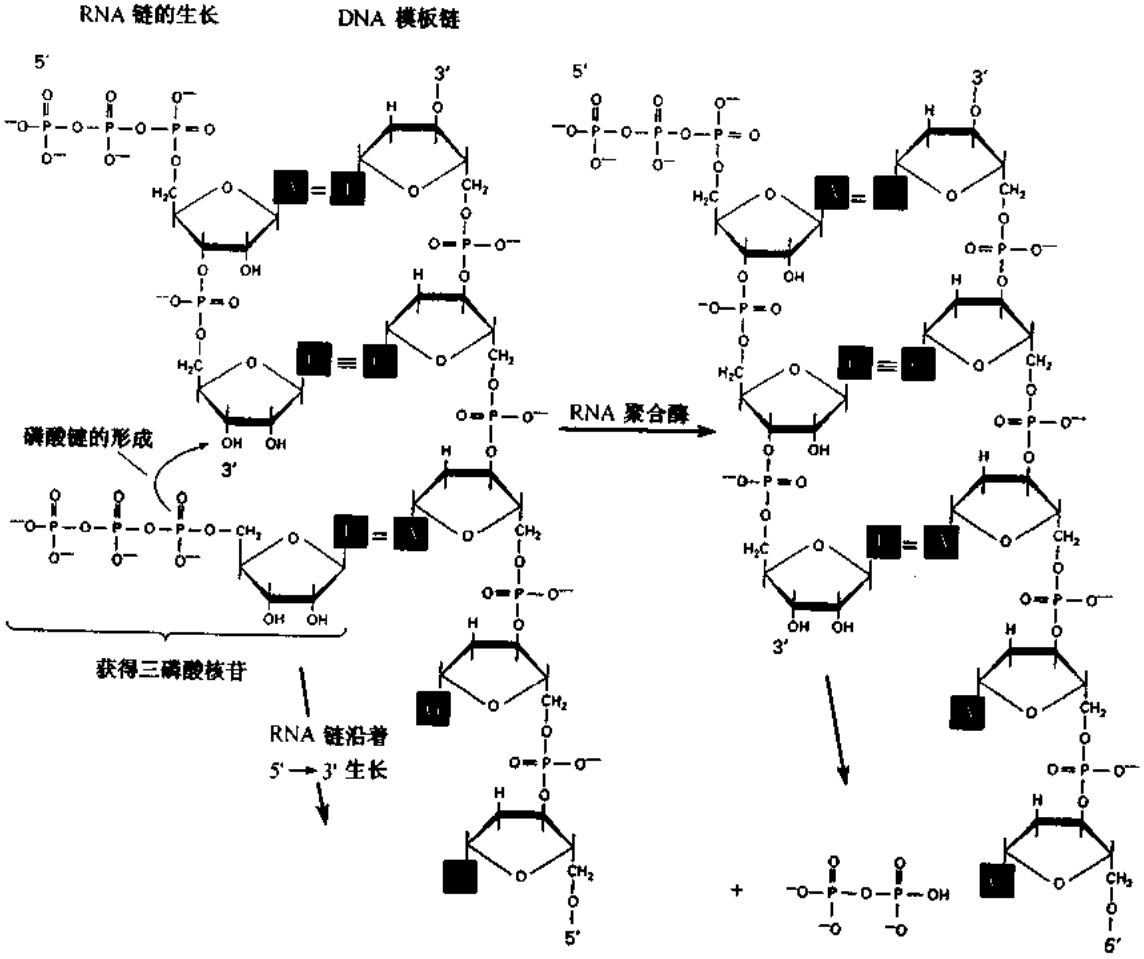


图 12－4　RNA 的酶促合成。(仿 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 12－5)

第三节　细菌基因的转录

一、细菌的 RNA 聚合酶

每种细菌都编码 RNA pol，其作用和 DNA pol 相似，它需要 DNA 模板，Mg 离子，和 4 种 3－磷酸核苷（ATP，UTP，CTP 和 GTP），缺少其中任何一种底物转录不能进行。RNA pol 和 DNA pol 也有 2 点不同：①RNA pol 没有任何校对功能；②RNA pol 能起始新的 RNA 链。

对细菌 RNA pol 深入的研究发现：*E.coli* 的 RNA pol 是一种复合酶，可以合成 3 类不同的 RNA(tRNA，mRNA 和 rRNA)，由 5 种多肽亚基组成（图 12－5），其大小和功能如表 12－1 所示：

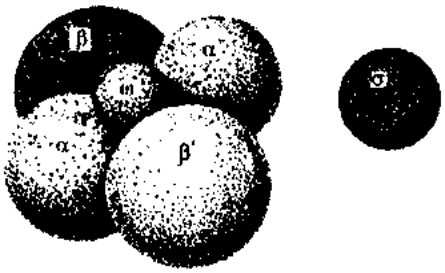


图 12－5　*E.coli* RNA 聚合酶的亚基组成。
（仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 12－6）

表 12－1　*E.coli* RNA Pol 的结构和功能

亚基	基因	分子量(kDa)	数目	组分	可能的功能
α	*rpoA*	40	2	核心酶	酶的连接，装配
β	*rpoB*	155	1	核心酶	和底物(核苷酸)结合
β′	*rpoC*	160	1	核心酶	模板结合
ω		10	1	核心酶	
σ	*rpoD*	70	1	σ因子	和启动子结合，识别模板链

参照 B. Lewin: Genes Ⅴ, 1994, Table 14.1, 14.2.

RNA pol 需执行多种功能：①识别 DNA 双链上的起始子；②使 DNA 变性，在启动子处解旋成单链；③ 通过阅读启动子序列，RNA pol 确定它自己的转录方向和模板链。④ 最后当它迅速到达终止子时，通过识别停止转录。

二、原核生物转录的起始延伸

1. 启动子的结构和转录起始

启动子(promoter)是位于基因 5′的调控区域(图 12－6)。基因表达时，RNA pol 识别这个区域中的保守序列并与之结合，启动转录的起始。原核生物不同基因的启动子虽然结构有一定的差异，但明显具有共同的特点。①结构典型，都含有识别(R)，结合(B)和起始(I)三个位点；②序列保守，如－35 序列，－10 序列结构都十分保守；③位置和距离都比较恒定；④直接和聚合酶相结合；⑤常和操纵子相邻；⑥都在其控制基因的 5′

端;⑦决定转录的启动和方向。

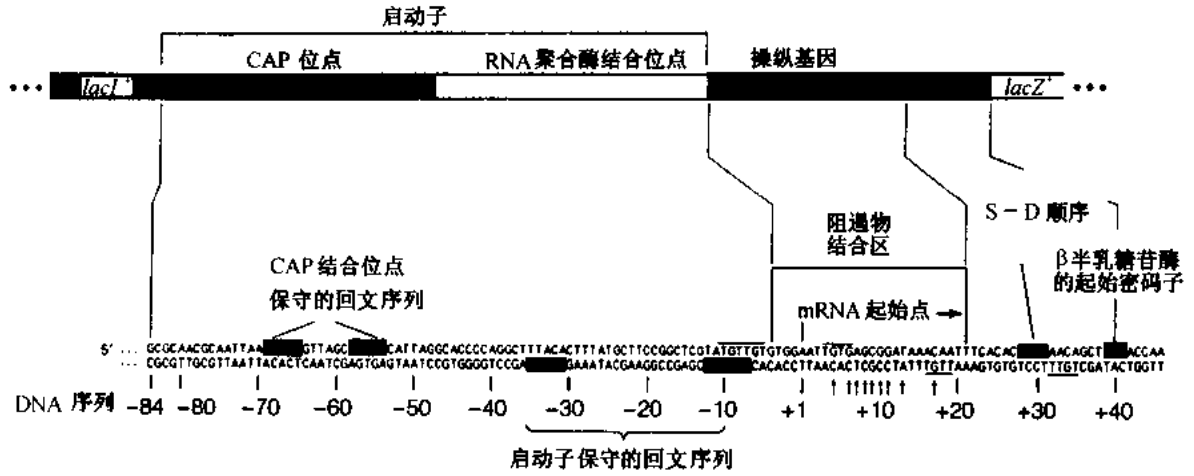


图 12－6 *E.coli* lac 基因的启动子和操纵子结构。(仿 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 16－13)

(1) －10 序列

－10 序列是由 Pribnow 和 Schaller(1975) 发现的，故也称为 Pribnow 盒 (Pribnow box)。其保守序列为 TATAAT($T_{80}A_{95}T_{45}A_{60}A_{50}T_{96}$)(图 12－7),位于－10bp 左右,其中 3′端的"T"十分保守。此序列 A.T 较丰富,易于解链。它和转录起始位点 I 一般相距 5bp。其功能是：①与 RNA pol 紧密结合;② 形成开放启动复合体;③使 RNA pol 定向转录。

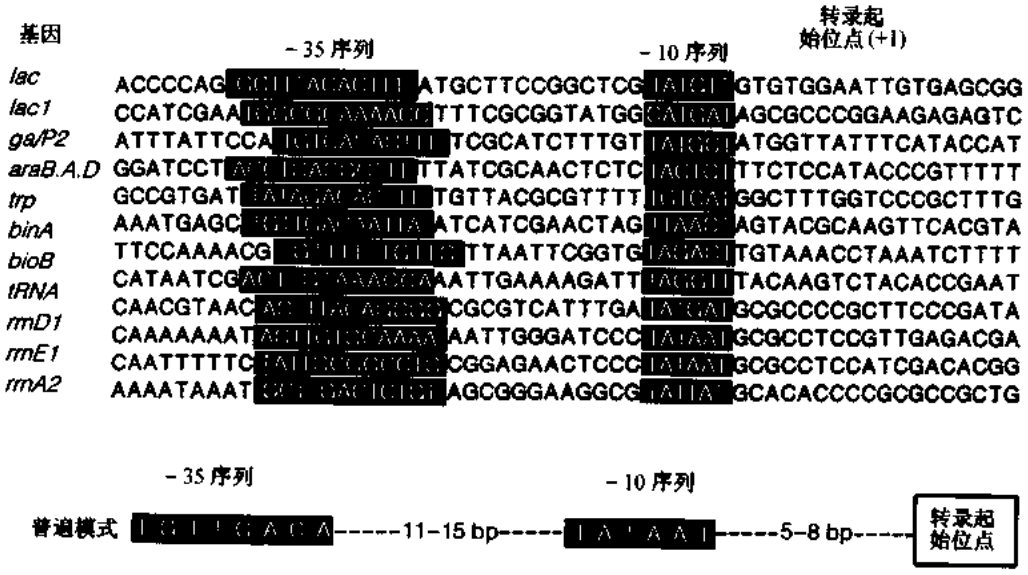


图 12－7 原核生物启动子的结构特点。(转引自 Griffiths, A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 13－8)

－10 序列的碱基组成对转录的效率影响很大，若－10 序列发生以下突变，TATAAT→AATAAT,转录效率会下降,故称为**下降突变**(down mutation)。若 *lacO* 的－10 序列发生以下突变,TATGTT→TATATT,转录效率会上升,故称为**上升突变**(up muta-

tion)。前者可能由于的$\overrightarrow{\underset{AT}{TA}}$堆集能要小于$\overrightarrow{\underset{TT}{AA}}$的堆积能，所以突变后双链的解链比突变前要多耗能量，所以影响转录效率，上升突变可能是由于$\overrightarrow{\underset{AC}{TG}} \rightarrow \overrightarrow{\underset{AT}{TA}}$不仅堆集能降低了，而且氢键也减少了，所以比突变前更易打开双链，转录效率也会提高。

(2) - 35 序列

- 35 序列又称为 **Sextama 盒** (Sextama box)，其保守序列为 TTGACA ($T_{82}T_{84}G_{78}A_{65}C_{54}A_{45}$)，与 - 10 序列相隔 16 - 19bp。其功能是：①为 RNA pol 的识别位点。RNA pol 的核心酶只能起到和模板结合和催化的功能，并不能识别 - 35 序列，只有 σ 亚基才能识别 - 35 序列，为转录选择模板链。② - 35 序列和 - 10 序列的距离是相当稳定的，过大或过小都会降低转录活性。这可能是与 RNA pol 本身的大小和空间结构有关，也就是说 RNA pol 的有关功能。

(3)转录起始位点(I)

一般都将转录开始时模板上的第一个碱基定成(+ 1)，在原核中常为 A 或 G，而且位置固定，不像真核基因，有时会发生前移的现象。

2. 转录的起始

起始的过程可分为四个阶段(图 12 - 8)。

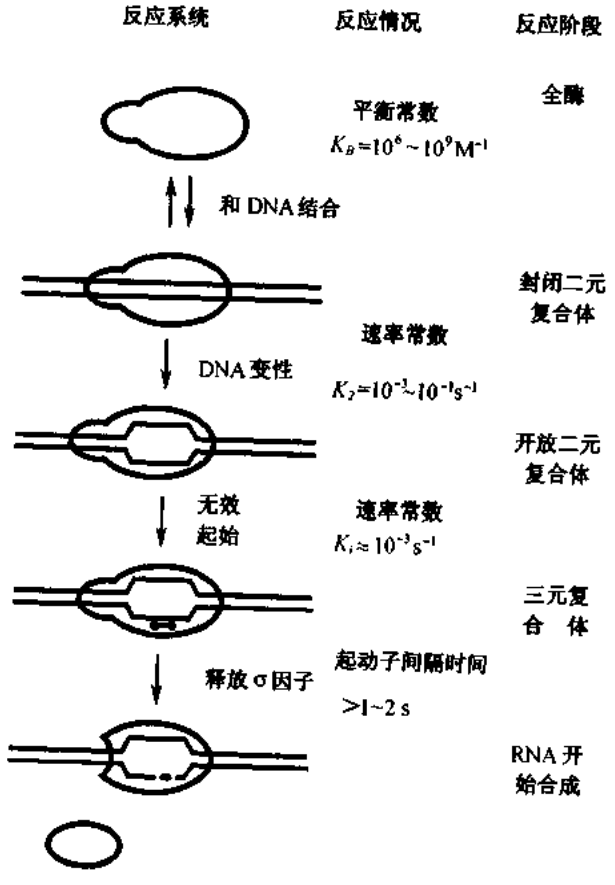


图 12 - 8 RNA 合成起始的几个阶段。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11 - 12)

(1) 核心酶在 σ 因子的参与下与模板的 DNA 接触，生成非专一的，不稳定的复合物在模板上移动；

(2) 起始识别：全酶与模板的启动子结合，产生封闭的**“酶－启动子二元复合物”**(closed binary complex)；

(3)全酶紧密地结合在启动子的 －10 序列处，模板 DNA 局部变性，形成**“开放的启动子二元复合体”**(open binary complex)；

(4)酶移动到转录起始点上，第一个 rNTP 转录开始，σ 因子释放，形成**酶－启动子－rNTP 三元复合体**。

全酶——启动子反应开始是通过形成封闭复合体产生一种松散的结合反应。由于这种封闭二元复合体的形成是可逆的，所以常以平衡常数 K_B 来描述它。K_B 的数值变化范围是宽的。

由于酶的结合，被结合的 DNA 中有一个很短的区域(－10 序列)熔解，使这个封闭复合体变成一个“开放复合体”(open complex)，此开放复合体被称为紧密结合(tight binding)反应。以上反应是不可逆的，所以用速率常数 K_2 来描述。$K_2 = 10^{-3} \sim 10^{-1} s^{-1}$，可见反应速度是快的。

下一步是头两个核苷酸的结合，然后在它们两之间形成磷酸二酯键。这样就产生了含有 RNA、DNA 和酶的**三元复合体**(ternary complex)。三元复合体的结构是通过速率常数 K_i 来描述的，K_i 要比速率常数 K_2 快得多。酶不用向前移动，核苷酸可以进一步加到 RNA 链上，直到增长 9 个碱基。当每个碱基加上后，酶可能释放 RNA 链。酶先和第一个碱基结合，然后产生无效起始 (abortive initiation)，一个无效起始循环通常合成一

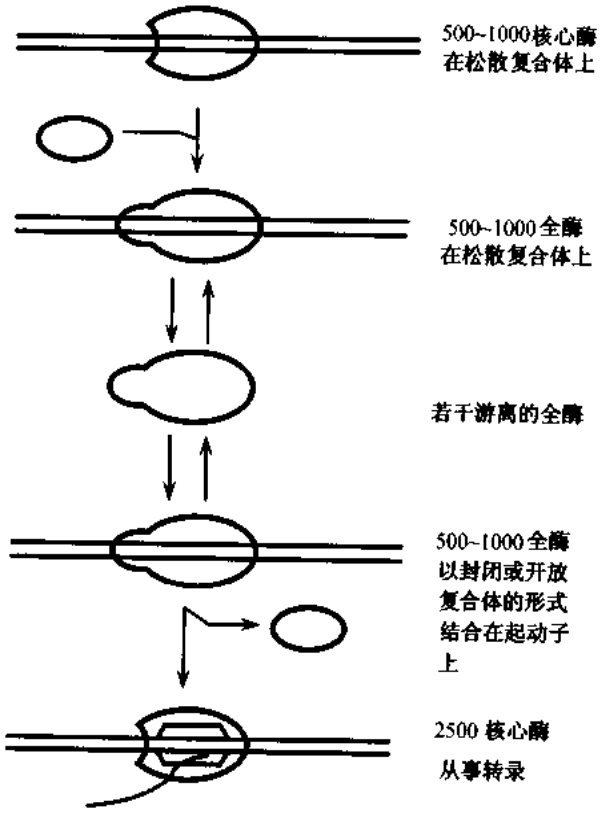


图 12－9　核心酶和全酶在 DNA 上的分布。(仿 B. Lewin: Genes V, 1997, Fig. 11－12)

个 2～9 碱基的寡核苷酸。

当起始成功后，酶就释放 σ 因子并形成一个由 RNA pol 的核心酶－DNA－新生 RNA 构成的"转录延伸三元复合体"。

起始反应可以用 3 个参数来描述，以下式表示：

$$R + P \underset{}{\overset{K_B}{\rightleftharpoons}} RP_C \xrightarrow{K_2} RP_0 \xrightarrow{K_i} RPi \cdot RNA$$

（R：RNA pol；P：启动子；RP_C：R－P 封闭复合体；RP_0：R－P 开放复合体；RPi·RNA：RNA pol－P－RNA 启动三元复合体）

这些特殊的参数表示聚合酶在启动子上停留多长时间，实际是启动子的间隔时间，它的极小值是 1～2 秒，即在 1～2 秒内建立起始最大频率，小于每秒一次。然后这个酶延着模板移动，而 RNA 链延伸第 10 个碱基。

RNA 核心酶和全酶在 DNA 上的分布如下（图 12－9）：①σ 因子和 1/3 的 RNA pol 结合成全酶分布于非特异位点的松散复合体和启动子上的二元复合体之中；② 其中半数的 RNA pol 的核心酶从事转录；③ 余下的核心酶大量存在于闭合松散复合体中。因为这些酶很快地参与复合体，而慢慢离开复合体，所以即使有游离的核心酶，也是非常少的；④有多少全酶是游离的呢？我们不知道，但估计其数量是很少的。

起始时当 RNA pol 全酶结合到 DNA 上，它从－55 一直延伸到＋20，覆盖了 77～80bp 形成起始复合物。在起始期它保持这个状态，然后发生起始区 DNA 的解链和 RNA 链的合成和延伸；很快 σ 因子就离开核心酶，使转录进入延伸阶段，RNA pol 从－55 收缩到－33，只覆盖 60bp 的 DNA 形成了开始的延伸复合体（图 12－10），因为－35 序列

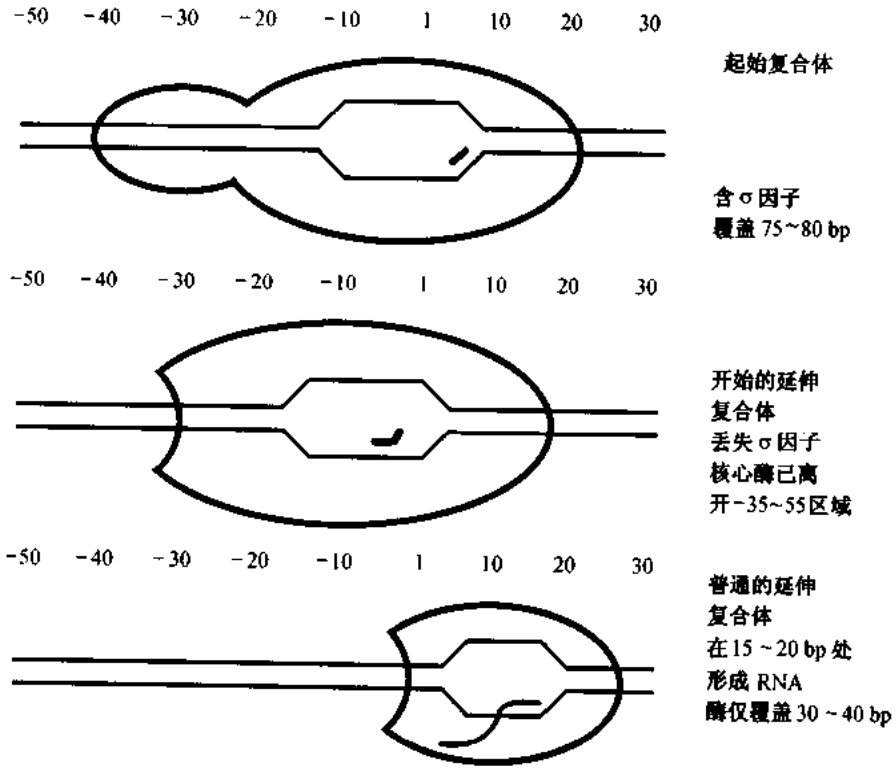


图 12－10　RNA 聚合酶起始转录。（仿 B. Lewin：Genes Ⅴ，1994，Fig. 14－8）

是起始时供 σ 因子识别时用，现已经起始，而且 σ 因子也已和核心酶分开，上游区域不必要了。当 RNA 链延伸到 15－20 碱基时，核心酶就向前移动。就形成普通的延伸复合体。此时核心酶只覆盖 30bp。目前还不知道这种变化是否涉及重要的酶的形态改变，也不知道是否为 DNA 解除一些紧张状态，(RNA 多聚酶长度为 160A，可覆盖松弛形 DNA 50bp)，这就意味着若和更长的 DNA 结合，必然涉及到 DNA 的弯曲。

3. 转录的延伸

传统的观点认为延伸是很单调的过程，即酶延着 DNA 向前移动一个 bp RNA 链也就伸长 1nt，其实这个过程是很复杂的(图 12－11)，上述仅是反应的平均状态。在延伸的时候，酶后端边缘的分界线也可作为 RNA 链延伸的末端。即酶的后端向前每移动一个 bp，RNA 延伸上也就加上了一个 rNTP，但酶的前端并没有移动，仍保持原来的位置，只不过酶整体收缩了一个 bp 的长度。酶内部它所覆盖的 DNA 双链的开放区及 RNA 的生长点(3′端)都向前移动了 1 个 bp。当 RNA 链已延伸到多个 nt 时，酶的前端突然向前一下子延伸 7～8bp。像是尺蠖向前运动的方式。图 12－11 的模型表示延伸时 RNA pol 以稳定收缩和突然伸展的方式在 DNA 上“爬行”，稳定的收缩是指 RNA pol 从 35bp 长连续地收缩到28bp 然后前端又突然向前伸展 8bp，RNA pol 又恢复到 35 bp 长。

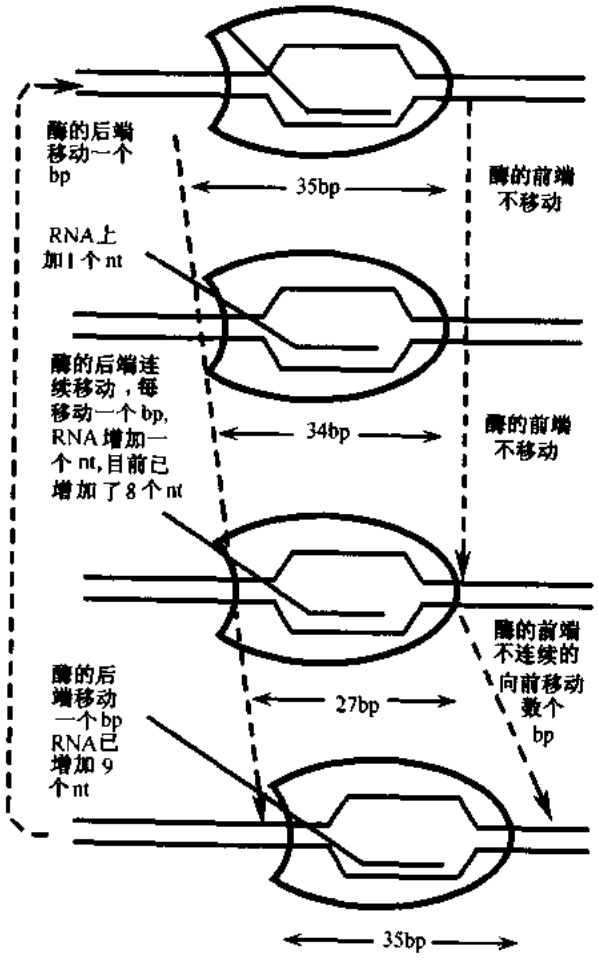


图 12－11　在延伸时 RNA pol 的移动特点。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11－7)

三、原核生物转录的终止

原核生物转录的终止处有特殊结构的存在，称为**终止子**(terminator，t)。RNA pol 能识别 t 位点在此处停止，然后释放 RNA，最终 RNA pol 也脱离模板，终止了转录。

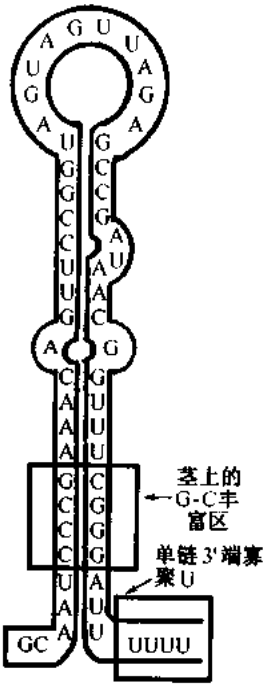


图 12－12　强终止子的结构。
(B. Lewin: Genes Ⅵ, 1997, Fig. 11－25)

在原核细胞中有两种不同的终止子，一种是**强终止子**，另一种是**弱终止子**。强终止子在体外实验中，无需其他任何因子的帮助就可以终止核心酶，这种终止子被称为**内部终止子**(intrinsic terminators)。弱终止子需要在一种蛋白质因子 ρ(rho factor)的帮助下才能终止，所以又称为 **ρ 依赖性终止子**(Rho-dependent terminator)。

强终止子的结构有三个特点（图 12－12)：①有回文结构存在。由它转录出的 mRNA 可形成茎环结构，可阻止 RNA pol 的前进；②茎的区域富内含 G－C，使茎环不易解开；③强终止子 3′端上有 6 个 U，由于它和模板形成的连续 U－A 配对较易打开，从而便于释放出 RNA。

有的转录单位在体外转录时并不发生终止，RNA pol 只是暂停在终止子上，但很快又恢复了 RNA 的合成。ρ 因子作为一种蛋白，若加到体外合成系统中就可使 RNA pol 在一定的位点终止，产生带有 3′端的 RNA 分子。这种作用称为 ρ－依赖性终止。ρ 是 *E. coli* 中的一种重要蛋白，在 *E. coli* 中种类很少，大部分 ρ 因子是在噬菌体的基因组中编码。

ρ 因子识别的序列早已被鉴别，这种序列若上游发生缺失，可能导致通读（read through)。此暗示 ρ 识别一个位于终止位点前的不连续区域，长 50～90nt。分析这段序列发现其共同特点是 C 丰富，G 缺乏(图 12－13)。

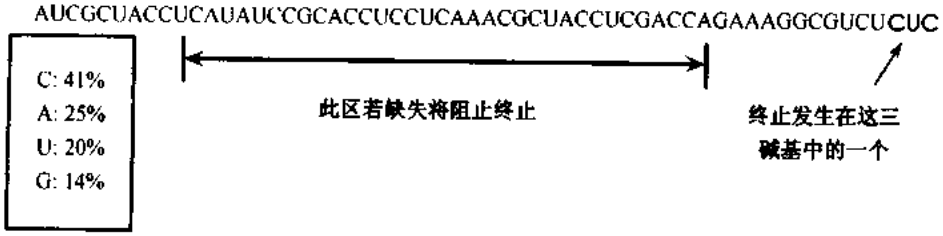


图 12－13　ρ 依赖性终止子，在终止位点前有一个富含 C 而少 G 的顺序。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11－26)

C：41%，A：25%，U：20%，G：14%。作为一般的规律，ρ－依赖性终止子的效率随着富 C/贫 G 区域长度的增加而增加。

还有一种 ρ 依赖性终止子，它也可形成茎环结构，但在茎中的 G·C 含量少，茎环易打开。在其 3′端也没有寡聚 U(图 12－14)，因此 RNA pol 合成此段 RNA 后，由于茎

环的存在可使聚合酶暂停，若此时ρ因子追上来和其结合，那么转录即可终止，若无ρ因子聚合酶还可越过终止子进行"通读"。

对于涉及真核 RNA pol 终止的信号因子和辅助因子知之甚微。各种聚合酶用不同的机制来终止。在真核中重现了原核生物 RNA 聚合酶内部终止子的同样特点：RNA 的二级结构和寡聚 U 尾巴对转录终止是重要的。

ρ因子在终止阶段可单独地显示其功能。它是 46kDa 的蛋白，作用时可能作为一种 6 聚体(275kDa)存在。其功能是作为 RNA pol 的一种辅助因子，当其浓度为 RNA pol 浓度的 10%时在体外可发挥最高的活性。

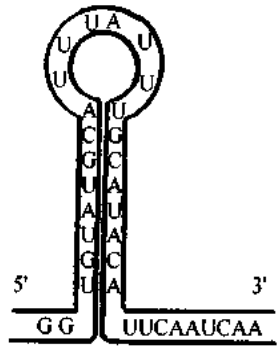


图 12-14 噬菌体人 T_RP 依赖。

ρ因子的作用是识别 DNA，RNA 还是 RNA pol 呢？ρ因子具有依赖 RNA 的 ATPase 活性和解旋酶活性，要求(50nt 的多核苷酸的存在，此表明ρ是结合 RNA。一个 6 聚体ρ因子作用在单个的 RNA 基质上。ρ因子的作用模型如图(12-15)所示，ρ结合到 RNA 链终止子上游的某一点，有可能需要一个特殊序列或某一类型的序列，结合可能发生

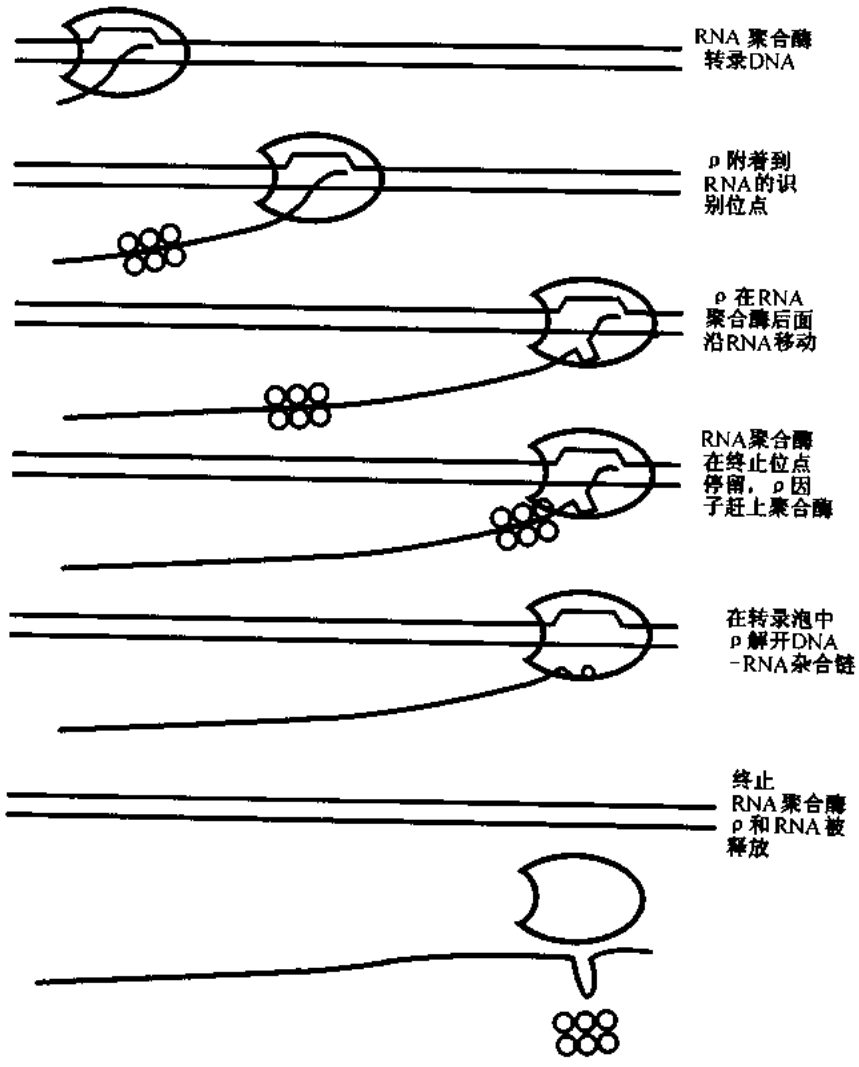


图 12-15 原核转录的ρ依赖性终止机制。(B. Lewin: Genes V, 1997, Fig. 11-27)

在 5′端，ρ 因子结合以后延着 RNA 向 3′端移动。

ρ 因子怎样追上 RNA pol 呢？一种可能性是 ρ 因子延着转录本移动比 RNA pol 延 DNA 移动要快。当 RNA pol 达到终止子时就暂时停了下来，那么在此位置 ρ 就会追上 RNA pol。

ρ 因子是通过直接作用于 DNA 和 RNA 的连接或通过间接作用 RNA pol 来释放 RNA 的呢？在没有 RNA pol 的情况下，ρ 因子可以导致 RNA－DNA 杂合链的分离，ATP 的水解用来提供反应的能量。

ρ 因子很可能直接在转录泡中接触 RNA－DNA 杂合链并导致其解链。或是由于解链的结果，或是由于 ρ 因子和 RNA pol 的相互作用，通过释放 ρ 和 RNA pol 使转录完全终止。

ρ 因子延着 RNA 的移动这一假设产生了一个关于转录和翻译两者之间关系的重要预言：ρ 因子首先要接触 RNA 上的结合序列，而且能延着 RNA 移动。正在翻译的核糖体可能会阻止其前进。当在翻译时发生了无义突变时 ρ 因子可以在终止子处和 RNA pol 接触(图 12－16)。

这个模型可以解释一些困惑的现象。有时，在转录单位中一个基因发生了无义突

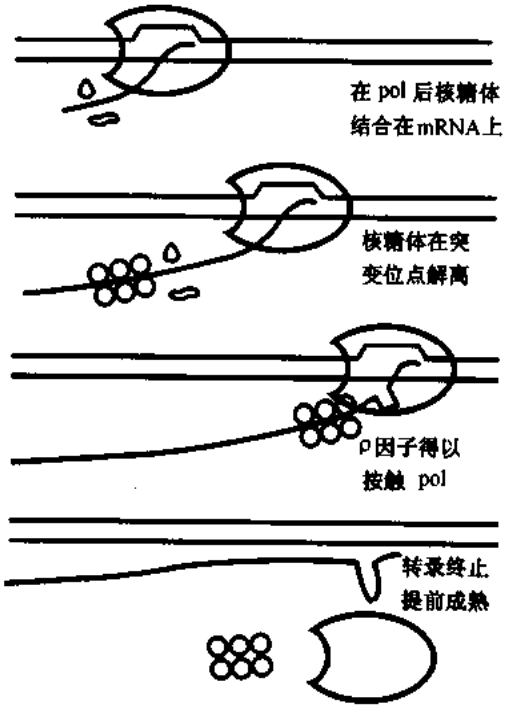


图 12－16　无义突变时 ρ 因子可能产生的终止作用。(B.Lewin: Genes V,1994.Fig.16－6)

变（nonsense mutation），将会阻止其后面基因的表达，这个作用称为极性（polarity）突变。共同的特点是缺乏与转录单位后面部分的相关 mRNA。

假设转录单位中存在 ρ 依赖性终止子，在正常情况下核糖体的阻隔，使 ρ 因子无法接触到 RNA 聚合酶，直到遇到正常的终止密码时核糖体解离，ρ 因子才有机会追上聚合酶。当前面的基因发生无义突变时，核糖体在无义突变拉点解离(图 12－16)，ρ 提前得以接触聚合酶，使转录终止，产生了部分 mRNA 释放出来。

整个的转录过程以及所涉及的酶和蛋白质因子，图 12－17 已可概括。

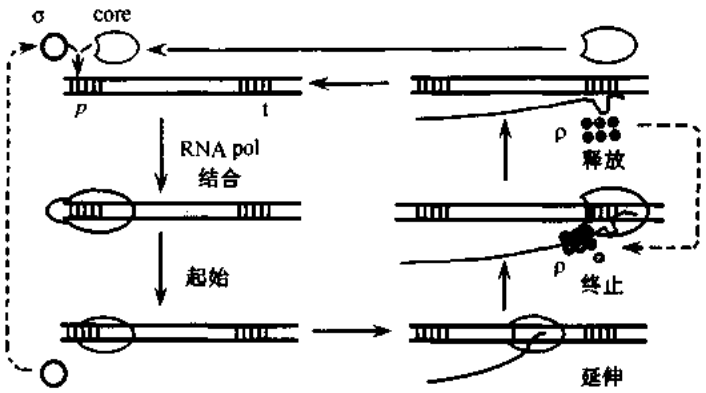


图 12-17 转录的过程。

四、RNA 聚合酶和噬菌体转录

在一般情况下，小的噬菌体是利用宿主细胞的 RNA pol 来进行转录。但大的噬菌体都有一个基因编码 RNA pol 来转录噬菌体的基因。如 T7 的基因 1 就专门编码 RNA 聚合酶。而宿主 *E.coli* 的 RNA 转录 T7 的部分染色体，包括基因 1。基因 1 的转录本在细菌的细胞中产生单链的多肽，这就是 T7 RNA 聚合酶。此酶再转录 T7 其余的噬菌体染色体。与此同时，其他基因的转录本转录产生蛋白抑制宿主 RNA 聚合酶的活性。所有这些酶的作用都是使噬菌体染色体的转录优于宿主染色体的转录。这是由于噬菌体的启动子和其 RNA Pol 有一种特别的亲密关系，而 *E.coli* RNA pol 的活性又被抑制。

T4 噬菌体在细菌的细胞中采用不同方式进行转录，不像 T7 噬菌体那样，而是有一个基因编码一种蛋白和 *E.coli* 的 RNA pol 酶相互作用，来改变它们的底物特异性。改变后的宿主聚合酶只转录噬菌体的基因而不能转录细菌的基因。

第四节 真核生物基因的转录

真核生物的转录过程和原核是相似的，所不同的地方在于：①原核只有一种 RNA 聚合酶而真核细胞有三种聚合酶；②启动子的结构特点不同，真核有三种不同的启动子以及其他与转录有关的元件；③真核的转录有很多蛋白质因子的介入。

一、真核的 RNA 聚合酶

1960～1961 年几乎同时由 S.Weiss，J.Hurwitz 和 Audrey Sterens 在 *E.coli* 及各种真核和原核细胞中发现了 RNA 聚合酶。

真核 RNA 聚合酶的特点如表 12-2 所列。每种酶涉及到不同类型的 RNA 合成。RNA pol Ⅰ是惟一在核仁中的 RNA 聚合酶，它催化合成两种大分子 rRNA(28S rRNAs 和 18S rRNAs)和 2 种小分子 rRNA 中的一种(5.8S rRNAs)，这些分子都在核糖体上。RNA pol Ⅱ只出现在核质中。它涉及 mRNA 和某些 snRNAs(核内小分子 RNA，small muclear RNA)，还有 RNA 的前体(hnRNA：heterogeneous muclear RNA，核内不均一 RNA)的合成。RNA pol Ⅲ也在核质中，它合成下列的 RNA：① tRNAs 将氨基酸运送到核糖

体上的 RNA；②5S rRNA 也是核糖体的一种组分；③某些 snRNA，这些 snRNA 涉及 RNA 的加工。

不同的真核聚合酶对各种抑制剂的敏感性也不同(表 12－3)。

表 12－2　真核生物的三种 RNA 聚合酶的特点

RNA pol	位置	产物	相对活性	对 α-鹅膏蕈的敏感性
pol Ⅰ	核仁	28S, 18S, 5.8S rRNAs	50%～70%	不敏感
pol Ⅱ	核质	hnRNA, mRNA, 某些 snRNAs	20%～40%	高度敏感
pol Ⅲ	核质	tRNA, 5S rRNA, 某些 snRNAs	～10%	片段特异，中等敏感

表 12－3　转录的抑制剂

抑制剂	靶酶	抑制作用
利福霉素	细菌全酶	和 β 亚基结合，抑制起始
链霉溶菌素	细菌核心酶	和 β 亚基结合，抑制起始
放射线素 D	真核 pol Ⅰ	和 DNA 结合，阻止延伸
α-鹅膏蕈	真核 pol Ⅱ	和 RNA pol Ⅱ 结合

引自 B. Lewin: Genes Ⅴ, 1994, Table 29.2.

RNA 合成抑制剂主要分两类，一类是模板结合抑制剂，另一类是聚合酶抑制剂，常用的原核及真核抑制如表 12－3。真核酶的主要特点是它们对双环十八肽的 α－鹅膏蕈（α－amanitin）的反应。在动、植物及昆虫的细胞中，RNA pol Ⅱ 的活性可被低浓度的 α－鹅膏蕈所抑制，但 α－鹅膏蕈却不抑制 pol Ⅰ。pol Ⅲ对 α－鹅膏蕈的反应，不同的生物有所差异。在动物细胞中高浓度的 α－鹅膏蕈可抑制转录，在昆虫中不受抑制。

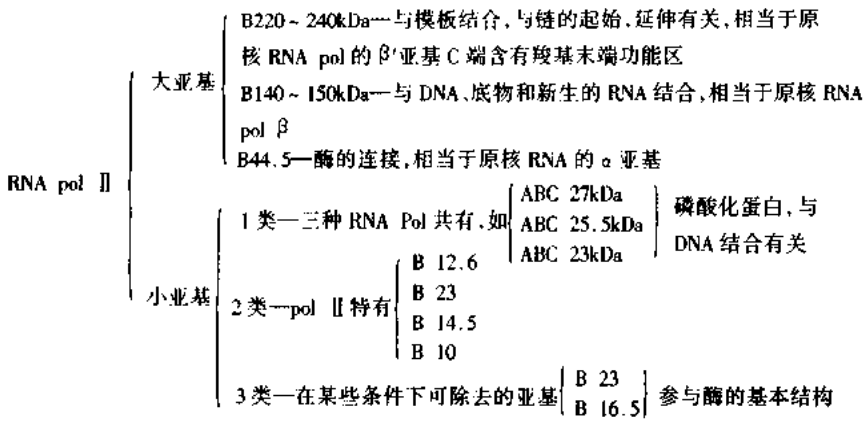
所有真核 RNA 聚合酶都是大分子蛋白质，分子量可达 500kDa 或更多。他们典型的有 8－14 个亚基。提纯的酶能依赖模板进行转录 RNA，却不能在启动子上选择起始，还没有一个真核的 RNA 聚合酶由提纯的亚基进行重建过，这就产生一个问题，所有的亚基是否都是重要的成分。在啤酒酵母中鉴定出只有 10～11 亚基是转录所必要的。

真核 RNA pol Ⅱ 的一般构成最典型的例子是啤酒酵母，RNA pol Ⅱ 有三个大亚基和细菌的 RNA 聚合酶同源，他们构成了基本的催化装置，具有特殊反应的能力。两个大亚基带有催化位点。余下的亚基有 3 种是所有真核 RNA pol 所共有的。大部分亚基是以单拷贝存在的，但有 3 种亚基有 2 个拷贝(表 12－4)。

RNA pol Ⅱ 的大亚基有一个**羧基末端功能区**（carboxy－terminal domain, CTD），它含有一个多次重复的一致序列 Tyr－Ser－Pro－Ser－Pro－Ser，这个序列是 RNA pol 独有的。在酵母中有 26 个重复，在哺乳动物中有～50 个重复。此重复数是重要的，因为如果切去其一半以上就会致死。此 CTD 在 Ser 或 Thr 残基上可高度磷酸化，这个酶带有非磷酸化亚基叫做 RNA pol Ⅱ a，而带磷酸化亚基叫做 RNA pol Ⅱ o，CTD 涉及起始反应较少。

激活线粒体和叶绿体转录的 RNA pol 较小，更类似于细菌的 RNA pol。当然细胞器的基因很小，RNA pol 只要转录少数几个基因，转录控制也很简单。因此这些酶和噬菌体酶相似，具有单一固定的功能而无需具有对更为复杂的环境反应的能力。

表 12－4 真核生物聚合酶的成分及功能



细胞器的 RNA pol 和噬菌体的 RNA pol 相似，因有单一固定的功能，分子量较小。

二、真核生物的启动子

由于真核生物中有三种不同的 RNA 聚合酶，因此也有三种不同的启动子，其中以启动子Ⅱ最为复杂，它和原核的启动子有很多不同：①有多种元件：TATA 框，GC 框，CATT 框，OCT 等；②结构不恒定。有的有多种框盒如组蛋白 H_2B；有的只有 TATA 框和 GC 框，如 SV40 早期转录蛋白，③它们的位置、序列、距离和方向都不完全相同(图 12－18)，④有的有远距离的调控元件存在，如增强子；⑤这些元件常常起到控制转录效率和选择起始位点的作用；⑥不直接和 RNA pol 结合。转录时先和其他转录激活因子相结合，再和聚合酶结合。

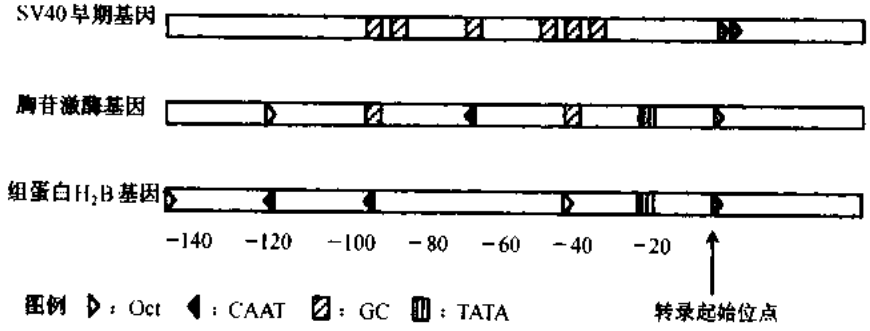


图 12－18 真核起动子含有的各种元件和分布。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 28－16)

1. Ⅱ类基因的启动子和调控区

Ⅱ类基因的启动子由核心元件和上游元件组成。核心元件包括 TATA 框和转录起始位点附近的启始子(initiator, Inr)。

在起始点一般没有同源序列，但 mRNA 的第一个碱基倾向 A，另一侧翼由 Py 组成（在原核启动子的 CAT 起始序列也有这种情况），称为起始子（initiator），一般由 $P_{Y2}CAP_{Y5}$ 构成，位于－3～＋5，可能提供 RNA pol Ⅱ识别。无论 TATA 是否存在，Inr

对于启动子的强度和起始位点的选择都是十分重要的。现已分离纯化了与 Inr 特异结合的蛋白质因子。

(1) 核心元件

TATA 框合又称 Hogness 框，Goldberg - Hogness 框，俚语称为金砖 (goldbrick)，其一致序列是：$T_{85}A_{97}T_{93}A_{85}A_{63}A_{83}A_{50}$，常在起始位点的上游 - 25 左右，相当于原核的 - 10 序列。但 - 10 是不可缺少的，而真核启动中也有缺乏 TATA 框的例子。其作用是：① 选择正确的转录起始位点，保证精确起始，故也称为**选择子** (selector)，当有的基因缺少 TATA 框时，可能由 Inr 来替代它的这一作用，如鼠的脱氨核苷转移酶(Tdt)基因就没有 TATA 框，但有 17bp 的 Inr；② 影响转录的速率。TATA 框的 8bp 的保守序列一般都是由 A.T 对组成，少数情况在其中的两个位点上由 G.C 对取代了 A.T，可见它较容易打开。当它的序列因发生突变和缺失而改变时就会影响它和酶的结合能力，从而影响转录的能力。在伴清蛋白基因中，当 TATA 框突变为 TAGA 后，转录效率大大降低。兔的珠蛋白基因当 TATA 框的保守序列 ATAAAA 人工突变为 ATGTAA 时转录效率会下降 80%。人的珠蛋白基因的 ATAAA 序列变为 ATGAAA 或 ATAG/CAA 时珠蛋白产量也会大大降低而出现地中海贫血症。

(2)上游启动子元件(UPE)

上游元件包括 CAAT box、GC box 等，它们的保守序列和结合的蛋白质因子也各

表 12 - 5　哺乳动物 RNA pol Ⅱ上游转录因子结合的元件

元件	保守序列	结合 DNA 的长度	蛋白质因子	大小(kDa)	丰度(每个细胞中)	分布
TATA box	TATAAAA	~ 10bp	TBP	27	?	普遍
CAAT box	GGCCAATCT	~ 22bp	CTF/NF1	60	30 万	普遍
GC box	同(表 12 - 4)GGGCGG	~ 20bp	SP1	105	6 万	普遍
Octamer	ATTTGCAT	~ 20bp	Oct - 1	76	?	普遍
Octamer	ATTTGCAT	23bp	Oct - 2	52	?	淋巴细胞
KB	GGGACTTTCC	~ 10bp	NFKB	44	?	淋巴细胞
KB	GGGACTTTCC	~ 10bp	H2 - TF1	?	?	普遍
ATF	GTGACGT	~ 20bp	ATF	?	?	普遍

引自 B. Lewin: Genes Ⅵ, 1997, Table 28.2

不相同。

CAAT box 的保守序列是 GG^{C}_{T} CAATCT，一般位于上游 - 75bp 左右紧靠 - 80，其功能是控制转录起始活性。能和 CTF(识别 CAAT 的转录因子)相结合。

GC box 的保守序列是 $^{C}_{T}$ $GGGCGG^{GGG}_{AAT}$，常以多拷贝形式存在 - 90 处。在真生物和病毒的一些启动子中常存在 GC 框，如鼠二氢酸叶酸还原酶启动子，猴基因组中的双向启动子，SV40，疱疹病毒等基因的启动子中(图 12 - 18)，可被转录因子 SP1 所识别。它的作用也是控制转录效率。

还有其他的一些 UPE，如八聚体(octamer)，KB，ATF 等(见表 12 - 5)。

(3)远端调控区

增强子　远端调控区较常见的是增强子 (enhancer)。至少在有时候增强子的存在可以增强启动子的转录活性。它是在 1981 年由 Benerji，Rusconi 小组和 Chambom 等发

现的，又称远上游序列(far upstream sequence)。其特点是：①具有远距离效应。常在上游－200bp 处，但可增强远处启动子的转录，即使相距十几 kb 也能发挥其作用；②无方向性。无论在靶基因的上游，下游或内部都可发挥增强转录的作用；③ 顺式调节。只调节位于同一染色体体上的靶基因，而对其他染色体上的基因无作用；④ 无物种和基因的特异性，可以接到异源基因上发挥作用，如将 SV40 的增强子接到兔 β－珠蛋白基因前，引入 LeLa 细胞，此珠蛋白基因转录增强 200 倍。⑤ 具有组织的特异性。 SV40 的增强子在 3T3 细胞中比多瘤病毒的增强子要弱，但在 Hela 细胞中 SV40 的增强子比多瘤病毒的要强 5 倍，抗体基因的增强子只有在 B 淋巴细胞中才起作用。增强子的效应需特定的蛋白质因子参与。⑥ 有相位性。其作用和 DNA 的构象有关。⑦ 有的增强子可以对外部信号产生反应。如热休克基因在高温下才表达。编码重金属蛋白的金属硫蛋白基因在镉和锌存在下才表达。某些增强子可以被固醇类激素所激活。

增强子的结构如图 12－19 所示。病毒 SV40 的增强子是最先被描述的，它由 2 个 72bp 的重复序列组成。位于上游－200bp 处。每个 72bp 的单元中都有 A、B 两个功能域。它们共含有 5 个序列元件(GT Ⅱ，GTI，Sph．11，sphI 和 P)，这 5 个元件对转录来说是很重要的。不同的序列模块是作为特殊转录因子识别序列。增强子的某些结构特征和启动子相似：如它们都是被反式作用因子(trans－acting factors)所识别的顺式作用序列(cis－acting seguence)，只不过增强子的作用距离比启动子长。

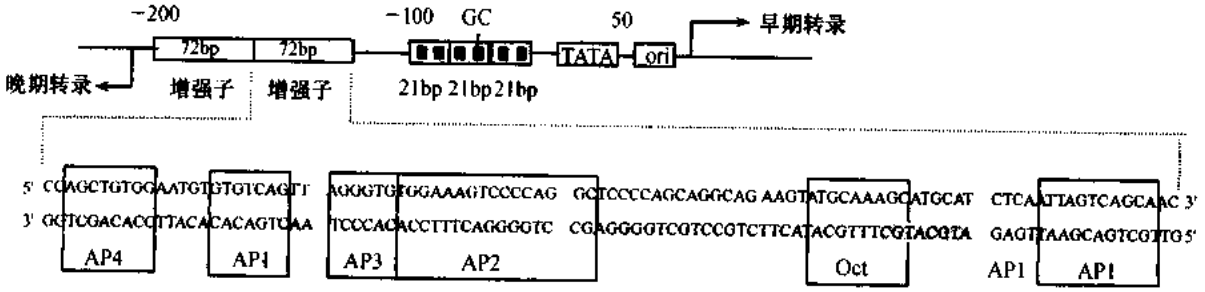


图 12－19 SV40 增强子的结构。AP 和 Oct 为调控蛋白的结合位点。

增强子为什么具有远距离作用呢？为解释这个问题前后曾有三种不同的假设：①拓扑效应；②滑动模型；③成环模型。拓扑效应说认为增强子的作用是诱导染色质结构变化，使核小体产生 DNase Ⅰ敏感区。SV40 的增强子区和免疫球蛋白基因的增强子区都发现存在 DNase Ⅰ的高敏感区。滑动模型增强子区域核小体常不能很好装配，结构松弛；在 SV40 的增强子区域存在左旋的 Z－DNA 区，也有助于双螺旋的打开，使和转录有关的蛋白质因子和 DNA 结合，然后再延着 DNA 滑向启动子。成环模型认为增强子通过一些蛋白质因子的介导可与远距离的启动子结合，使 DNA 形成了一个环，从而促进远距离的启动子的转录。这一模型已得到普遍的认可。一些实验的证据表明免疫球蛋白 K 链基因的增强子可以作用两个距离不同的启动子 (分别为 440bp 和 2.7kb)，而作用是相同的。若是按滑动模型，对距离的启动子的作用应高于远距离的启动子。但事实并非如此。从消耗能量的角度来说，长距离的滑动对生物是不利的，在进化中会受到选择的压力。成环模型也符合染色体的侧环模型和核基质的调控理论(见第 17 章) 也就是说 DNA 的特殊序列可以和核基质结合形成侧环，在某些细胞中某些基因通过环的形成使增强子区和启动子区相互靠近，使这些基因得以表达，显示了基因表达的组织特异性。看来环的形成主要是两种因素：①某些蛋白质因子的介导；②和核基质特异的结合(图 12－20)。

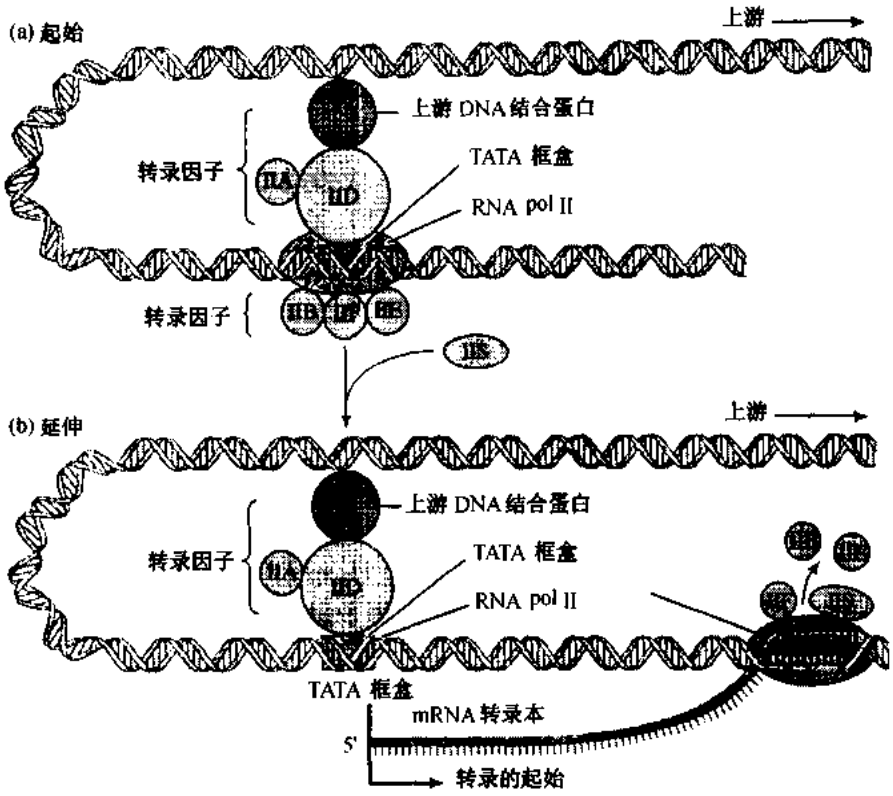


图 12－20　真核 RNA pol Ⅱ启动子和上游调控区成环连接，启动转录。
（仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 12－12）

减弱子　在某些基因的上游远端或下游远端具有负调节序列，其作用不受距离和方向的影响，叫做**减弱子**（dehancer）。如 c-*mos* 基因上游 0.8 或 1.8kb 处有一序列，使 c-*mos* 不易被反转录病毒的长末端重复序列（LTR）所激活。在 c-*myc* 基因的 3′端也存在着减弱子。

静息子　在酵母交配型转换的盒式模型中，左右两个沉默框（HML 和 HMR 见 17, 23 章）都不表达，此是由于在这两个框盒上游 1.5kb 处都有一个 E 片段，它可和阻遏沉默框盒表达的蛋白 SIR（1－4）（silent mating tyne information regulation）结合，起抑制作用，故称**静息子**（silencer），可作用 2.5kb 远的启动子。

上游激活序列（upstream activating sequences UASs）　UASs 是酵母中远上游序列类似于增强子。它仅影响转录程度，对位点选择不起作用。和增强子不同的是它有方向性，不能在启动子的下游起作用。它结合的转录因子是 GCN4 和 GAL4，识别位点为 ATGACTCAT。

2. RNA pol Ⅱ的转录起始

第一步是原称为 TF Ⅱ D 的转录启动因子和 TATA 框结合形成复合物，为 RNA pol Ⅱ指明结合位置。现在已知 TF Ⅱ D 含有两种蛋白，一种是识别 TATA 框的 TBP（TATA-binding protein），为 30kDa 的小蛋白。另一些亚基称为 TAFs（TBP 结合因子，TBP-associated factors）。有的 TAFs 是和 TBP 等当量的。而另一些 TAFs 存在量较少。很可能 TF Ⅱ Ds 含有不同的 TAFs 来识别不同的启动子。

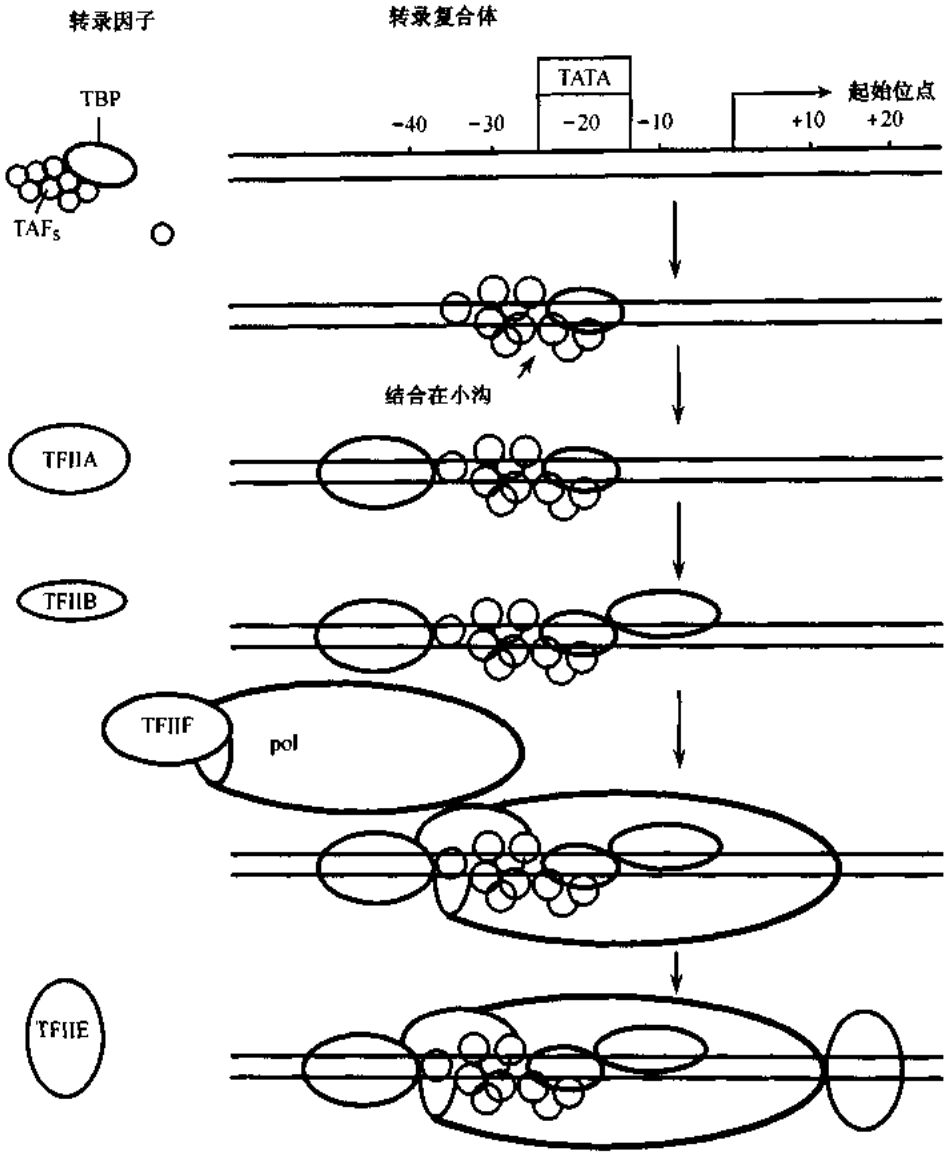


图 12-21 真核生物 RNA pol Ⅱ的转录起始。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 28-11)

这种观点认为 TBP 和不同的 TAFs 结合可以使其他的聚合酶识别启动子。TF Ⅲ B 是用于内部 pol Ⅲ启动子，SLi 是用于 pol Ⅰ启动子，但这两者都被看成是和 TBP 结的一种特殊 TAFs。TBP 分子并不能识别不同的启动子，它必须依赖于不同的亚基来识别特殊的启动子。TBP 必须具备在各种启动子上和这种因子或聚合酶相互作用结合的能力。在酵母中 TBP 基因突变会影响到不同类型启动子的转录。

TBP 作为第一个和 DNA 接触的因子十分引人注目，有是它被看成是一种“束缚”因子(commitment factor)，其实是和 RNA pol Ⅱ结合，使启动子转录，起到介导的作用。由于 TATA 框离转录起始位点的距离是固定的，识别它对 RNA 聚合酶来说是重要的。TBP 在启动上对各种聚合酶所起的一个相同的作用是将它们组人转录复合体中。

TF Ⅱ A 含有几个亚基(在酵母中有 2 个，在哺乳动物中有 3 个)，当它加入复合体中时，TF Ⅱ D 所保护的 DNA 区域就延伸更上游的区域。它可能通过解除 TAFs 的阻遏作用而激活 TBP，它的参与使 TF Ⅱ D 和 TATA 框结合得更稳定(表 12-6)。

表 12－6　RNA pol Ⅱ启动子的转录因子的结构和功能

因子	分子量	功能
RNAPol Ⅱ	≥10k	依赖模板合成 RNA
TF Ⅱ A	12，19，35k	稳定 TF Ⅱ D 和 DNA 的结合，激活 TBP 亚基
TF Ⅱ B	33k	结合模板链（－10～＋10），起始 Pol Ⅱ结合，和 TF Ⅱ E/F 相互作用
TF Ⅱ D	（TBP，30k）	TBP 亚基识别 TATA，将聚合酶组入复合体中，TAFs 识别特殊启动子
TF Ⅱ E	34k(β)，57k(α)	结合在 pol Ⅱ的前部，使复合体的保护区延伸到下游
TF Ⅱ F	38，74k	大亚基具解旋酶活性（RAP74），小亚基和 pol Ⅱ结合，介导其加入复合体
TF Ⅱ H		具激酶活性，可以磷酸化 pol Ⅱ C 端的 CTD，使 pol Ⅱ逸出，延伸
TF Ⅱ I	120k	识别 Inr，起始 TF Ⅱ F/D 结合
TF Ⅱ J		在 TF Ⅱ F 后加入复合体，不改变 DNA 的结合方式
TF Ⅱ S		RNA 合成延伸

TF Ⅱ B 分子量为 33kDa，它在起始点邻近区域，从－10 到＋10 可以部分地保护模板链，在 TATA 框下游与 DNA 松散结合形成复合物。它和 DNA 双链结合是不对称的，它还可以使 TF Ⅱ E/Ⅱ F 相互作用。

TF Ⅱ F 含有 2 个亚基，大亚基 RAP74，具有 AT－依赖性 DNA 解旋酶活性，它和起始时 DNA 链的熔解有关 。它的小亚基是 RAP38，和细菌的 σ 因子有部分同源，紧密地和 RNA pol Ⅱ结合。介导 pol Ⅱ和其他蛋白结合转录成复合体。TF Ⅱ F－pol Ⅱ和复合体连接时，与 TF Ⅱ B 的相互作用是重要的。

RNA pol Ⅱ结合位点在模板链上达到下游＋15，在非模板链上达到＋20，其长度贯穿了整个复合体 ，上游也被其保护了。

TF Ⅱ E 结合在 pol 的前部，使复合体的保护区延伸到下游双链的＋30 处。

TF Ⅱ H 和 TF Ⅱ J 在 TF Ⅱ F 之后也加入复合体中，但并不改变 DNA 的结合方式。TF Ⅱ H 具有激酶活性，可以磷酸化 RNA pol Ⅱ尾巴上的 CTD，CTD 的磷酸化使 pol Ⅱ从转录因子中释放出来，这样就能离开启动子开始延伸。

这一起始过程和细菌 RNA pol 催化的反应是相似的。RNA pol 的结合产生了闭合复合体。在最后阶段双链打开产生开放复合体。这种转变是和下游位点复合体影响的增加有关，多半是和 TF Ⅱ因子的释放有关。起始反应的双链分离阶段需 ATP 的水解，可能要释放某些转录因子。在基本的因子和 pol Ⅱ的相互作用中看来 TBP 和酶尾部的 CTD 之间的作用也是重要的。

在没有 TATA 框的启动子上起始如何发生呢？这种起始同样需要一些基本的转录因子，其中也包括 TF Ⅱ D。TF Ⅱ D 可能带有一个或多个 TAFs 来直接识别 Inr，并与之结合。现在还不清楚在和 TATA 结合中 TF Ⅱ D 怎样发挥其作用。Inr 序列可能作为一种位置元件，让转录因子结合于其上，且锚定这种复合体。TF Ⅱ D 看来必须以某种其他方式进入复合体而不结合在 TATA 框上。在这些启动子上 TBP 的功能很像在 pol 的启动子上和 pol Ⅲ的内部启动子上。

很多的基本因子含有多种亚基，所以多数分子量是很大的，可能涉及到约 20 多种肽，总分子量达到 500kDa，pol Ⅱ约有 10 个亚基分子量也有约 500kDa，看来 RNA pol 的起始涉及到非常大的复合体。

pol Ⅱ起始复合体提供了一个和原核转录有趣的对照。细菌的 RNA pol 很容易和 DNA 结合；σ 因子对于起始是必要的，但无助于延伸，因起始后它就被释放出来了。真

核的 pol Ⅱ只有在起始因子和 DNA 结合后才和启动子结合。这些因子的作用类似 σ 因子，使聚合酶识别启动子上的特殊序列，但已进化得比较独立。

3. RNA pol Ⅰ启动子

RNA pol Ⅰ的启动子变化最小，RNA pol 仅从单一类型的启动子转录 rRNA 基因，转录本含有大 rRNAs 和小 rRNAs 的序列，经过加工再将它们释放出来。

描述得最清楚的是人类细胞中的这种启动子，它由两部分序列构成，一是**核心启动子**(core promoter)或核心元件(core element)，位于起始位点的前后，从 -45 到 +20，负责转录的起始。另一部分是**上游控制元件**(upstream, control element UCE)，它从 -180 延伸到 -107，此区可增加核心元件的转录起始的效率。这两个区域都有一个特殊的成分，就是 G.C 丰富区(G.C 含量达 85%)。

RNA pol Ⅰ需要 2 种辅助因子：UBF1(上游结合因子 1)结合在核心启动子相关的特异序列中和 UCE 上。SL1 因子，它本身对这种启动子来说并非是特异的，但一旦 UBF1 和 DNA 结合了，那么 SL1 就可以协同结合在 DNA 上。当这两个因子都结合上了，RNA 聚合酶才能和核心启动子结合起始转录。很可能是这些因子结合在核心启动子上直接与 RNA pol Ⅰ相互作用，但还不清楚在 UCE 上的因子如何刺激核心启动子的起始。在 RNA pol Ⅱ启动子中远距离作用是一个重要的特点，研究得较为深入，但不知

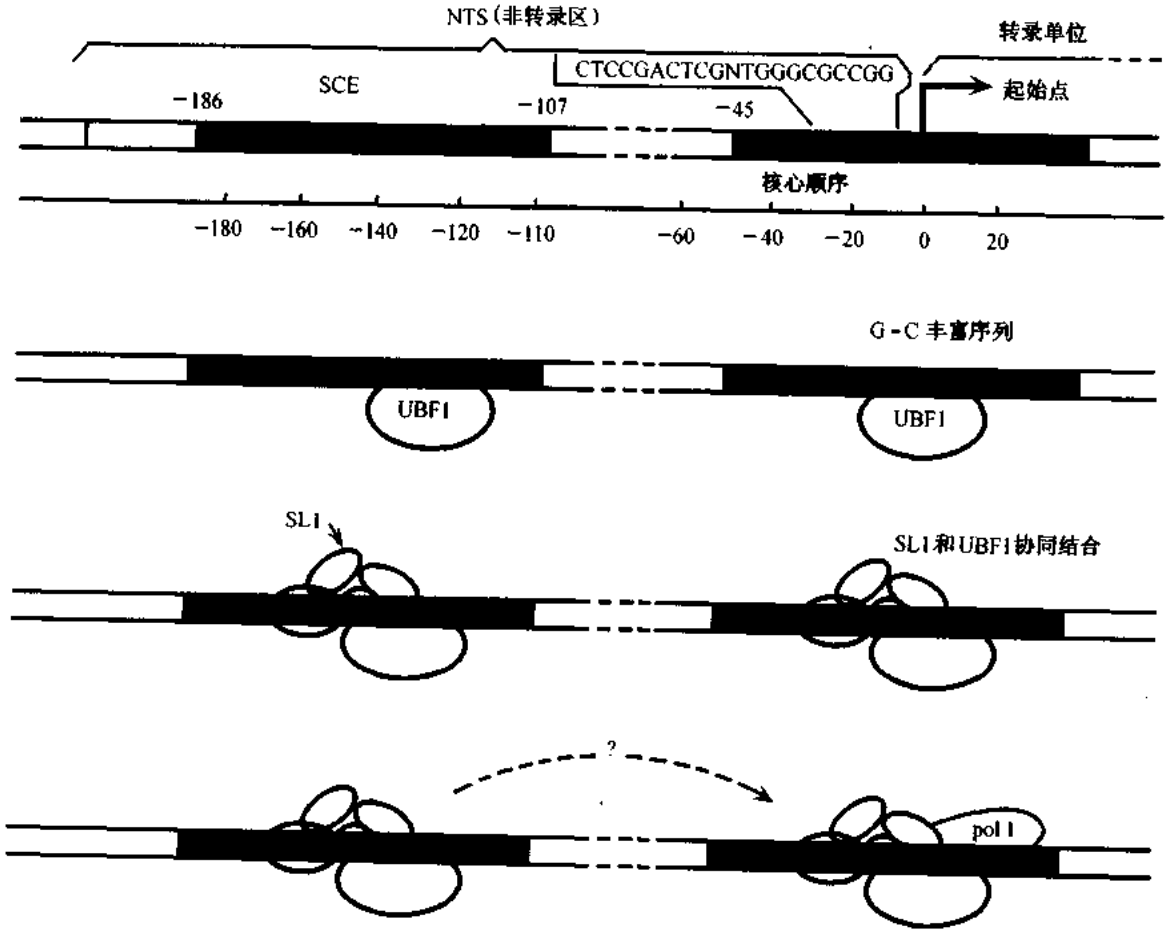


图 12-22 真核 RNA pol Ⅰ启动子的结构和起始图。(仿 B.Lewin: Genes Ⅴ, 1997, Fig.28-4)

道 RNA pol Ⅰ是否具有相同的机制。

UBF1 是一个单链多肽，它可以和核心区 UCE 的 G.C 丰富区结合（图 12－22）。UBF1 和 RNA pol 可在异源模板上发挥功能，如鼠的这种蛋白质因子和 pol Ⅰ能识别人类的基因。但 SL1 在起始转录时具有种的特异性。

SL1 含有 4 个蛋白，其中之一称 TBP，也是 pol Ⅱ和 pol Ⅲ起始时所需的一种蛋白质因子。可能在聚合酶起始后不久就发挥作用。TBP 在种间是保守的，但似乎并不具有种的特性，也并不一定要特异地和 G.C 丰富区结合。因此可能和 DNA 的结合的种的特异性是 SL1 另一些组分的功能。TBP 可能和 RNA pol 相互作用，和 RNA pol 的一个共同的亚基或一个功能域相结合。

SL1 的行为与细菌的因子相似，作为复合体的一个成员起作用。若分开来的话，它单独是不能和启动子特异结合的。但和其他组分结合起来就可以结合在启动子的特异区域。其初步功能可能是保证 RNA 聚合酶能位于起始位点的合适位置上。简而言之其功能是使 TBP 和其他蛋白质结合，为 pol Ⅱ和 pol Ⅲ提供转录因子。三种聚合酶起始的共同特点是依赖 TBP 和蛋白质结合构成的一种“位置”因子，这种位置因子对不同类型的启动子来说是特异的。

非洲爪蟾的 pol Ⅰ启动子并不分为两个区域，它从 +1 到 －141，其序列具备高度的种的特性，RNA pol Ⅰ在体外只能从密切相关的转录系统中转录 rRNA 基因。种的特异性是由转录因子决定的。

4. pol Ⅲ 启动子的结构及起始

人们通过转录起始区不同位点的缺失或突变来研究各种启动子的结构和功能。RNA pol Ⅲ启动子一般分成基因内启动子和基因外启动子两类（图 12－23）。它们是由不同的转录因子以不同的方法来识别的。5S RNA 和 tRNA 的启动子都属于 RNA pol Ⅲ启动子，但这些启动子比较特殊，位于起始位点的下游的转录区内，因此也称为**下游启动子**（downwtream promoter）或**基因内启动子**（intragenenic promoter）或称为内**部控制区**（internal contron regin ，ICR）。snRNA 基因的启动子和常见的启动子一样位于起始位点的上游，在以上两种不同的情况中都有特定序列存在，这些序列由转录因子识别而并不直接和 RNA 聚合酶结合。正是这些构成了特殊的 pol Ⅲ 启动子。

在非洲爪蟾的 5S rRNA 基因启动子在未鉴别出前，人们都以为它在起始位点的上游，缺失分析表明，当上游序列全部被切除后 5S rRNA 的产物仍照样合成。当缺失在基因内发生时仍可转录，只不过产物也缺少一段。但缺失发生在 +55 区域时，就不再转录。因此可以推测启动子位于起始点下游 +55，可能由于 pol Ⅲ起始转录要有一个固定的距离。当缺失发生在基因末端，但保留 80bp 序列时转录不受影响。但当缺失发生在 80bp 区域转录就会停止，表明启动子的下游分界域在 +80。

因此 5SRNA 的启动子是在基因中的 +55～ +80 之间。含有这个区域的 DNA 片段就可在其上游起始转录，起始点在 +55 的更上游的位置。

RNA pol Ⅲ 的三种类型启动子的结构如图 12－23 所示。用突变分析表明有两种类型的内部启动子（图 12－24），每一种都由两部分构成。这是两个短的序列元件中间间隔着一个可变化的序列。比较一下这两种类型的内部启动子表明：Ⅰ型内部启动子含有两个分开的 boxA（ T G G C N N A G T － G G ）和 boxC（C G G T

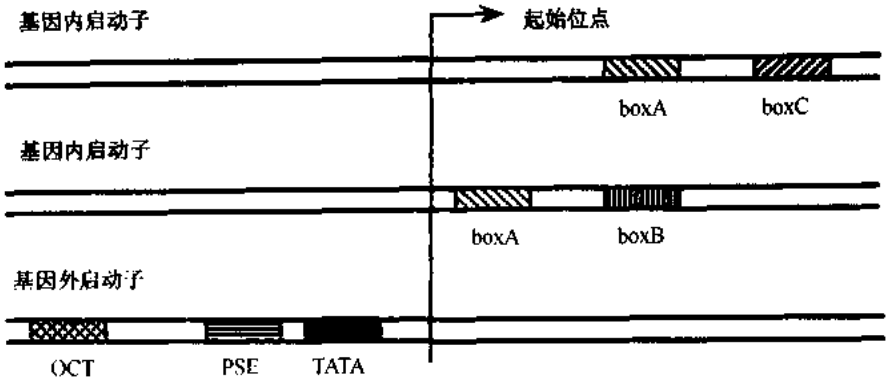


图 12－23 真核 RNA pol Ⅲ的基因内启动子和基因外启动子。(仿 B.Lewin：Genes Ⅵ，1997，Fig.28－6)

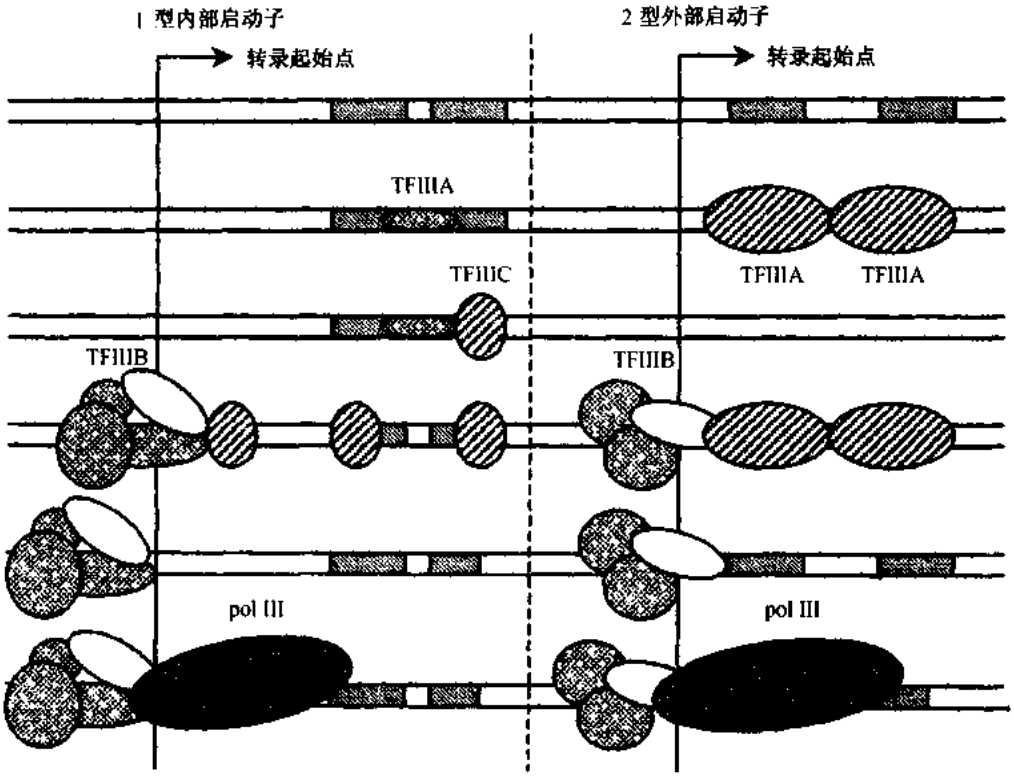


图 12－24 真核 RNA pol Ⅲ的下游启动子的结构和转录起始(仿 B.Lewin：Genes Ⅵ，1997，Fig.28－7)

C G A N N C C)序列。而Ⅱ型内部启动子含有两个分开的 boxA 和 boxB。Ⅱ型内部启动子中 boxA 和 boxB 之间的距离较宽。通常有功能的此类启动子中的两个 box 就不能紧紧连在一起。

图 12－24 表示内部启动子中的起始的各个阶段，这需涉及 3 个起始因子，TF Ⅲ A 已被克隆，是一种含 9 个锌指的蛋白(Zinc finger protein)(表 12－7)。TF Ⅲ B 含有 TBP 和两个其他的蛋白。TF Ⅲ C 是一种大分子蛋白复合物，至少有 5 个亚基，分为两

表 12-7　RNA pol Ⅲ启动子的转录因子的结构和功能

因子	结构	功能
TF Ⅲ A	38kDa,有 9 个锌指	结合于 1 型内部启动子(5S RNA 基因)的 C 框,使Ⅲ C 结合在 C 框下游,辅助Ⅲ B 定位结合
TF Ⅲ B	含 TBP 和另外二种蛋白	定位因子,使 pol 结合在起始位点上
TF Ⅲ C	含 τA 和 τB,有 5 个亚基	τB 结合Ⅱ型内部启动子(tRNA 基因)的 B 框,起增强子的作用。τA 结合 A 框,起启动子的作用;辅助Ⅲ B 定位结合
TBP	是Ⅲ B,Ⅱ D,SL1 的亚基	和特异 DNA 序列及 RNA pol 结合,使 pol 结合在正确的位点上
TⅡD	含 TBP 亚基	结合 TATA 框,确定选择 polⅢ
PBP	次近端结合蛋白	可能和Ⅱ D 一道辅助Ⅲ B 定位结合的另一途径

个功能域 τA(300kDa)和 τB,分别和 2 型内部启动子(tRNA 基因启动子)的 A 框和 B 框结合。B 框相当于增强子的作用,A 框相当于启动子的作用。在Ⅰ型内部启动子中(5S RNA 基因启动子)TF Ⅲ A 结合在 C 框上,使 TF Ⅲ C 结合在 C 框下游。在Ⅱ型内部启动子中 TF Ⅲ C 的结合使 TF Ⅲ B 依次结合在起始位点的近上游(图 12-23)。TF Ⅲ B 结合在起始位点上并和 TF Ⅲ C 相连。

这些因子作用的突出特点是当将 TF Ⅲ A 和 Ⅲ C 从启动子上除去(在体外用高浓度的盐)不影响起始反应。只要 TF Ⅲ B 仍结合在起始点的附近就能使 RNA pol 正确地结合在转录起始点。因此只有 TF Ⅲ B 才是 pol Ⅲ 所必需的起始因子。TF Ⅲ A 和 TF Ⅲ C 仅是一种**装配因子**(assembly factor),它们的作用是辅助 TF Ⅲ B 结合到正确的位置上。这样就解释了下游启动子怎样使聚合酶结合到上游位点。

TF Ⅲ B 的功能是作为一个"**定位因子**"(positioning factor)负责 RNA pol 结合正确位置上。就像 pol Ⅰ中的 SL1 那样。功能类似于原核的 σ 因子。它本身缺乏和 DNA 的结合能力,但可以和结合在 DNA 上的其他蛋白结合。前面已说过 TF Ⅲ B 含有 TBP,而 SL1 也含有 TBP,它可能作为 TF Ⅲ B 的一个亚基,直接和 RNA 相互作用。

内启动子虽然赋于这些基因转录的能力,但起始点对转录也具有一定的影响。若改变起始点上游区域就会改变转录效率。看来内部启动子的作用是提供识别;起始位点的作用是控制转录效率。

在第三种 pol Ⅲ启动子(图 12-23)有 3 个上游元件,这些元件仅在 snRNA 启动子中被发现,有的 snRNA 是由 pol Ⅱ转录,有的是由 pol Ⅲ转录。这些上游元件在一定程度上和 pol Ⅱ及 pol Ⅱ的启动子相似。TATA 元件看来和特异的聚合酶结合。

pol Ⅲ上游转录起始发生在起始点上游的一个很短的区域中,且含有 TATA 框。**次近端序列元件**(proximal sequence element, PSE)和八聚体 OCT 元件的存在大大增加了转录效率,结合在这些元件上的转录因子相互协同作用。

TATA 元件是供 TBP 识别的,TBP 亚基本身识别 DNA 序列,结合的其他蛋白有的可和 pol Ⅲ结合,有的对 pol Ⅱ特异,这就可以解释为什么 pol Ⅲ和这些启动子特异结合。TBP 及其结合蛋白的功能是使 pol Ⅲ正确地结合在起始位点上。TF Ⅱ D 也含有 TBP 可和 TATA 结合,促使 pol Ⅱ定位于起始位点附近。

TBP 是 RNA pol Ⅱ起始因子中的重要成分,它是定位因子的关键成分,通过不同机制结合在各种类型的启动子上。无论哪一种启动子,TBP 都是起始复合物的组成部分,与 RNA pol 的相互作用中发挥共同的作用。

在 RNA pol Ⅲ启动子上的这些因子的功能中有一种潜在的协同效应。TF Ⅲ B 在

启动子上结合在 pol 的前方，本身并不能和 DNA 结合。需要依靠Ⅲ A、Ⅲ C 协助，形成一种前起始复合体(preinitiation complex)结合在特异位点上，然后再直接和 pol 结合。启动子与 RNA pol Ⅲ识别的这种机制显示出起始点的建立对转录起始是十分重要的。

RNA pol Ⅲ 对任何特殊序列都没有内在的亲和性，它正好和位于起始点上游的一些起始因子相结合。使其正确定位。在 1 型和 2 型内部启动子中，含有 TBP 亚基的 TF Ⅲ B 就是起到这种作用，提供了位置信息。对于 pol Ⅲ 的上游启动子，转录因子识别特异位点形成含有 TBP 的复合体(图 12－25)，供 pol Ⅲ识别。这样不论启动子特异序列的位置如何，为了和 RNA pol Ⅲ的直接结合，它们总是有序地结合在起始点附近。

三、真核基因转录的终止

真核生物的 RNA 大部分都要经过加工，包括剪切和加上 poly(A)，因此对于其终止的情况了解很少，只有少数的例子搞得比较清楚。

前体 RNA 转录的终止随着组织的变化而变化。例如在爪蟾中有两个重要的转录终止位点：T2 和 T3。T2 是 28SrRNA 序列 3′末端下游约 235bp 序列。T3 是排列在下一个转录单位的上游约 200bp 处 (图 12－26)。T2 和 T3 都含有 7 个碱基的保守序列：5′－GACTTGC－3′。T2 是前体 rRNA 转录的初始终止位点。T3 是作为"失败－保险"的终止位点，由于 rDNA 转录单位是串联排列的，所以"失败－保险"终止系统让 RNA pol 分子可能在一个转录单位上起始和转录。

富含 A.T 区以及 RNA 的二级结构可能对终止也是重要的，还可能有终止因子存在。

有的 mRNA 并没有 poly(A)，其 3′端的形成不同于一般剪切或加尾形成的 3′端。这类 mRNA 突出的成员是编码组蛋白的 mRNA。其 3′端的形成依赖于二级结构。RNA 的终止子在茎环结构中，若发生突变阻止茎的形成也就阻止了 RNA 末端的形成。二级结构突变若使茎环恢复(并不要求原来的序列)形成末端的行为又得到恢复。这表明二级

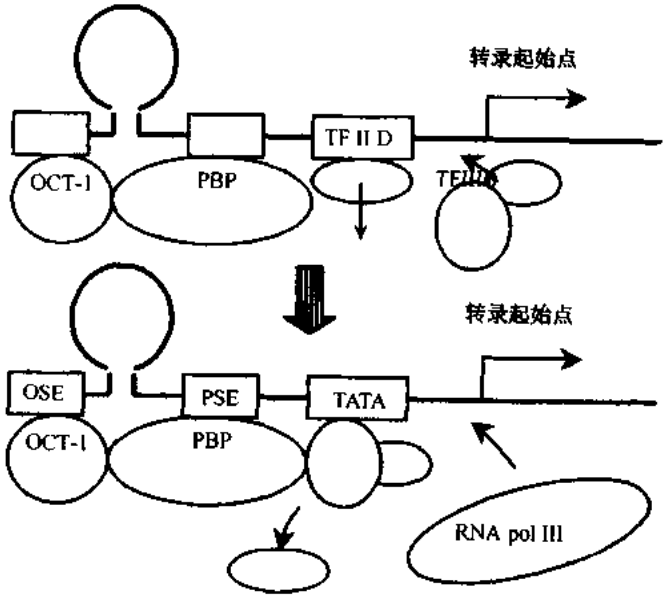


图 12－25 真核 RNA pol Ⅲ的上游启动子的转录起始。

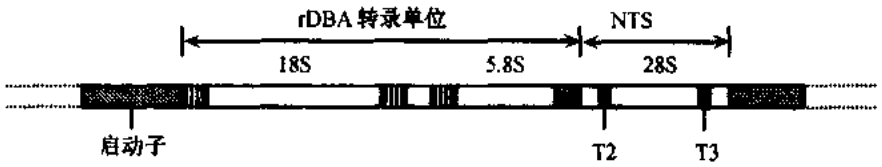


图 12 - 26　rDNA 的转录终止子。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 12 - 19)

结构的形成比严格的序列更为重要。

爪蟾组蛋白 H3 基因的 mRNA 在体外系统中产生可靠的 3′末端，无 poly(A)尾巴。当将一个克隆的海胆 H3 基因注入爪蟾卵母细胞时，很快启动转录，但终止发生在可变的位点上。但将海胆核的提取物和此基因一道注入，转录的 mRNA 有固定的 3′末端。

这个核内的提取物有 3 种活性成分，①热不稳定因子；②结合在剪切位点上游发夹结构上的因子；③称为 U7 snRNA 的 56nt 的 RNA，其 5′端和 3′端剪切位点的一个短的序列可以互补。

这个反应是在组蛋白 H3 mRNA 和 U7 snRNA 之间进行（图 12 - 27），上游的发夹结构和下游与 U7 snRNA 配对的序列在各物种的组蛋白 mRNAs 中是保守的。U7 snRNA 有一段序列可能和组蛋白 mRNA 的保守序列配对，而 3′端有一个大的发夹结

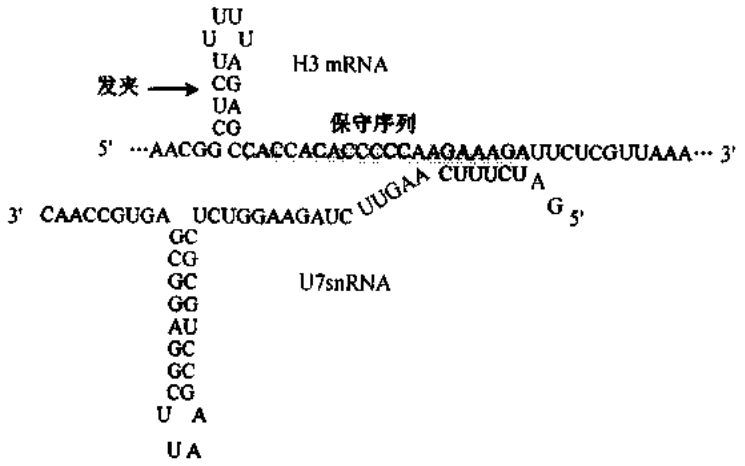


图 12 - 27　H3 mRNA 3′端的产生依赖于发夹结构和与 U7 snRNA 配对的保守序列。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 30 - 29)

构，U7 snRNA 的这一序列和发夹结构在人和海胆中都保留着。

组蛋白下游保守序列的突变减低了与 U7 snRNA 配对的能力，从而抑制了 3′端的加工。U7 snRNA 上若发生补偿性突变，恢复了和组蛋白 3′的互补性。则也恢复了 3′端的加工能力。这表明 U7 snRNA 的功能是和组蛋白 mRNA 碱基配对。我们还不知道在此阶段中 snRNA 在终止和剪切过程中的步骤。

第十三章

转录后加工

基本概念

编码蛋白质基因的转录本是线性的 RNA，其长度的范围宽广。原核 mRNAs 转录后很少加工，而真核 mRNAs 要在 5′端加"冠"，3′端加上 poly(A)尾巴等修饰。

大部分真核 mRNA 含有内含子，内含子一般不编码氨基酸。初始转录本经加工切除掉内含子，产生成熟的有功能的 mRNA 分子。内含子的切除发生在前体 mRNA 和核内小分子核糖核蛋白(snRNPs)的复合物中。这个复合物称为剪切体(spliceosome)。snRNPs 中含有 snRNAs 和各种蛋白。

tRNA 分子将氨基酸带到核糖上，在核糖体上聚合成蛋白链。各种 tRNA 长度相同，含有一些修饰碱基和相似的三维结构。

tRNA 的前体分子 5′端含有前导序列(Leader)和 3′端拖尾(trailer sequences)序列，加工时酶将这些序列切除掉。有的真核前体－tRNA 含有内含子，在加工阶段也要被切掉，此加工步骤和 mRNA 的加工步骤是不同的。

核糖体是细胞器，在核糖体上蛋白质进行合成。原核和真核的核糖体都含有 2 个大小不同的亚基。每个亚基是由一个或多个 rRNA 分子与许多核糖体蛋白构成。真核的核糖体比原核核糖体大而复杂。

真核的 18S，5.8S 和 28S rRNA 是由 rDNA 转录成单个的前体－RNA 分子。此前体 rRNA 在分子的末端和 rRNA 序列之间含有间隔序列 (spacer seguences)。当前体 rRNA 加工时间隔序列也会被切除掉，产生成熟 rRNAs 和有功能的核糖体亚基。

在有的前体 rRNA 中有内含子存在。这些 RNA 折叠成二级结构，加工时可自我剪接，这些内含子称Ⅰ类内含子，自我剪接时不需要任何其他蛋白。

转录后的加工 (posttranscriptional modification) 是指将各种前体 RNA 分子加工成成熟的各种 RNA。在原核和真核生物中 mRNA，tRNA 和 rRNA 都是来自于转录，在真核中还有第四种 RNA 分子即 sn RNA，同样也是转录的产物。

加工(processing)的结果不外乎 3 种形式：①**减少部分片段**：如切除 5′端前导序列，3′端尾巴和中部的内含子；②**增加部分片段**：5′加帽，3′加 poly(A)归巢和通过编辑加入一些碱基；③**修饰**：对某些碱基进行甲基化等。近年来在内含子的剪接、归巢和编辑等方面有不少新的发现和进展，现将加工的过程分述如下。

第一节　tRNA 和 rRNA 的加工

一、原核的 tRNA 和 rRNA 的加工

在翻译过程中参与蛋白质合成的 tRNA 和 rRNA 都不是最初的转录产物，这是由于：①它们的 5′端都是单磷酸，而原始的转录产物 5′应是三磷酸；②分子比初始转录物小；③tRNA 含有特殊的碱基，这些碱基只有通过一些化学修饰才可以得到。

1. tRNA 的加工

原核的 tRNA 初始转录本多为**多顺反子**(polycistron)，也就是几个 tRNA 分子串连在一起。这有三种不同的情况：①串联的 tRNA 分子都是相同的，如在 27′的 $tRNA^{Tyr}$ – $tRNA^{Tyr}$；②串联的 tRNA 分子是不同的，如 71′的 $tRNA^{Ile}$ – $tRNA^{Ala}$ – $tRNA^{Thr}$；③由 tRNA 和 rRNA 串联组成。少数的 tRNA 前体为**单顺反子**(monocistron)如 43′的 $tRNA^{Ser}$(图 13 – 1)。tRNA 的加工分成 3 个阶段(图 13 – 2)。

(1)“斩头”，形成 5′末端。RNase P 是一种不常用的酶，是由蛋白和 RNA 组成的复合体。其 RNA 长 375nt，分子量为 130kDa。根据 RNase 42℃突变株的研究表明 RNase P

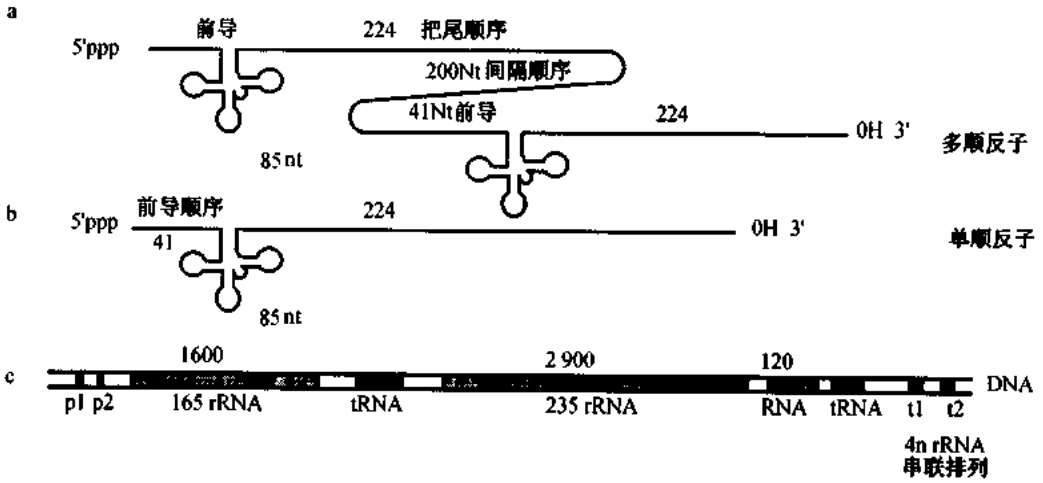


图 13 – 1　三种 tRNA 前体分子。

具有内切酶的活性，可切除 *E. coli* 前体 tRNA 5′端的**前导序列**(41nt)，此酶不识别特殊的序列，而识别二级结构——发夹所组成的 tRNA。这就是 S. Altman **提出的外部引导理论**。处在底物内的高级结构区，可供 RNase 识别，最终仍存在于成熟产物中，此 tRNA 的高级结构就称为外部引导区。

(2)去尾，形成 3′ – OH 末端。此过程由内切酶和外切酶的共同参与。前者识别发夹结构，后者识别 CCA 序列，但还不知道是哪一种核酸内切酶具体起作用。修整 3′端的外切酶是 RNase D，在 CCA 末端它一次切除(或逐步切除)3′的碱基。这是对于具有 CCA 末端的 1 型 tRNA 前体分子而言，对于没有 CCA 序列的 2 型 tRNA 前体分子来说当然不可是通过 RNase D 来修整，具体是何种 RNase 外切酶也还不清楚。RNase Ⅱ具有外切酶活性，但它一旦介入 tRNA 的加工，就有可能将整个的 tRNA 序列降解掉。这样它可能只参与一般的降解，而不是参与 tRNA 的加工。

在细菌中两种类型的前体 tRNA 的 3′序列是不同的。1 型分子有 CCA 尾巴，它作为 3′修复的信号和最后的界线。但二型分子并没有 CCA 序列，所以修整后需要加上 CCA，现在也不知道Ⅱ型分子加 CCA 的末端是怎样产生的，是像 I 型分子那样由 RNase D 来切，还是由其他的酶来作用都不清楚。在真核中所有前体分子都是Ⅱ型的。加 CCA 是由 tRNA 核苷转移酶(tRNA nucleotidyl transferase)来完成的，加上 CCA 的 tRNA 分子才成为有活性的 tRNA 分子。此反应的过程是

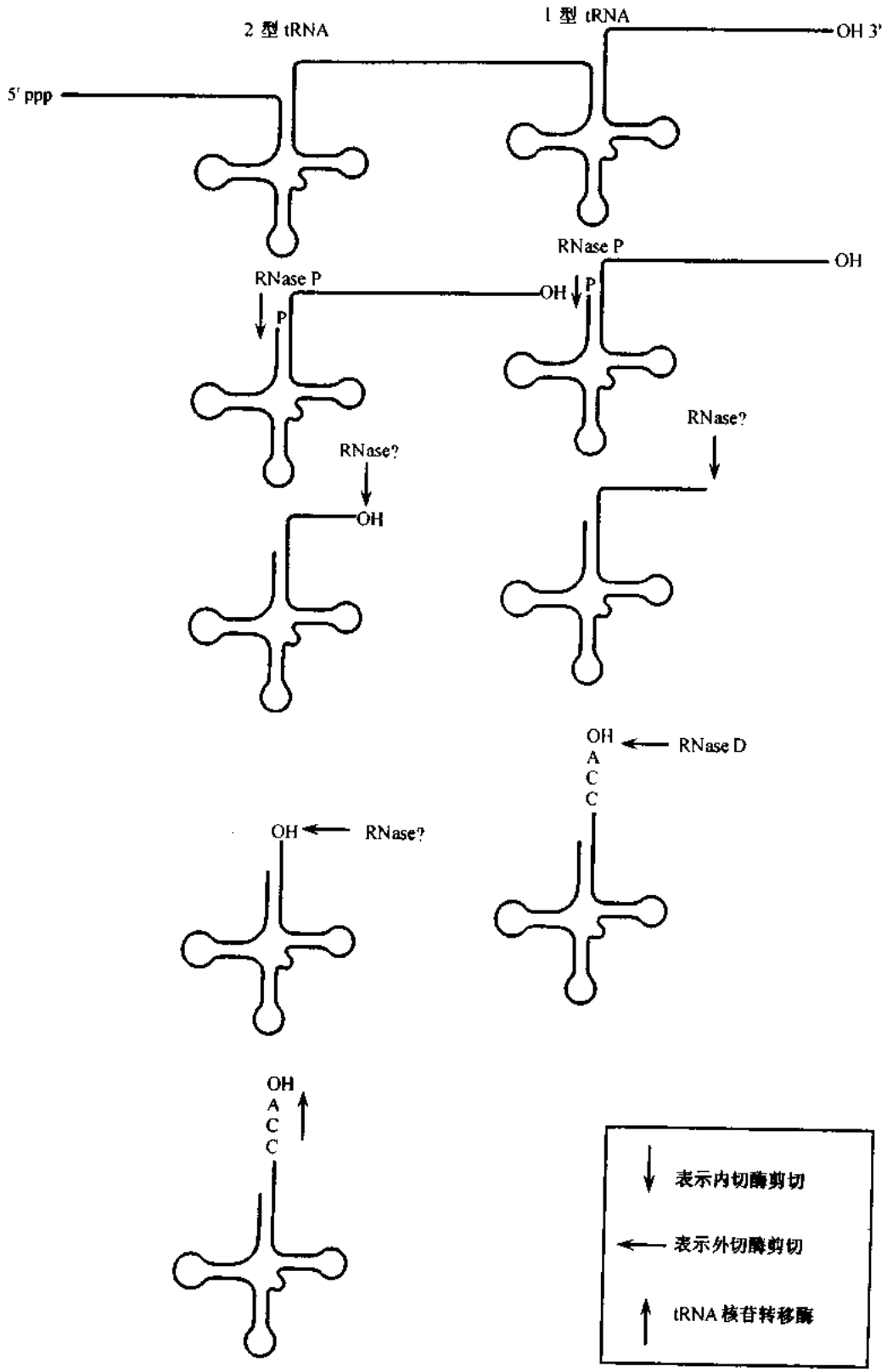


图 13－2　原核生物 tRNA 前体的剪切。(引自 B. Lewin: Genes Ⅵ, 1997, Fig. 14－17)

$$\text{缺-CCA 的 tRNA} \xrightarrow[\text{tRNA 核酸转移酶}]{\text{2CTP: 1ATP}} \text{tRNA-CCA} + \text{PPi}$$

(3)修饰。在前体的一些专一部位的碱基需要通过甲基化酶，硫醇酶，假尿嘧啶核苷化酶等的作用进行修饰成为特殊的碱基，如氨基酸臂上 5′的 4-硫尿苷(4tU)，D 臂上的 2 甲基鸟苷（2mG），TψC 臂上的假尿苷（ψ）以及反应密码子环上的 2 异戊腺苷(2ipA)(图 13-3)。

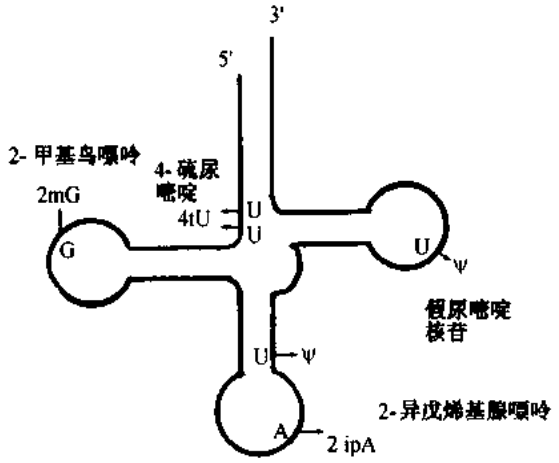


图 13-3　tRNA 前体中特殊碱基的化学修饰。

2. rRNA 的加工

在 *E. coli* 中 rRNA 有 7 个转录单位，名为 *rrnA-G*，它们在染色体上并不紧密连锁。rRNA 序列是保守的，但还不知道这些序列的同源性是怎样产生的。每个转录单位都含有 16S、23S、5S rRNA 及一个或几个 tRNA。但 tRNA 的数量，种类及位置都不固定，或在 16S rRNA 和 23S rRNA 之间的间隔序列中，或在 3′端 5S rRNA 之后。每个转录单位中含有等比例的 16S、23S、5S rRNA 是很有意思的，因为它们仅存在于核糖体中且是等比例的，因此串联转录单位保障了它们的等量关系。

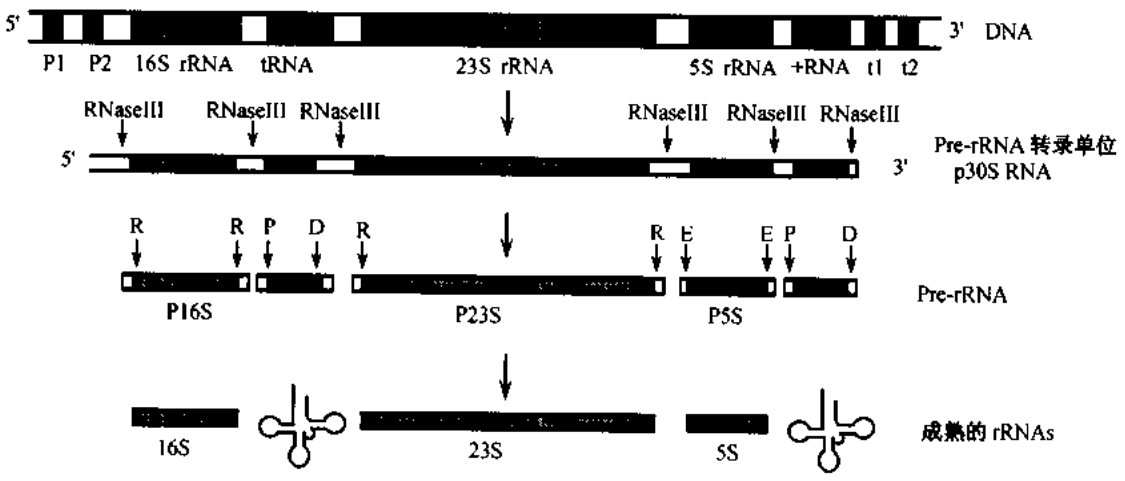


图 13-4　原核生物 rRNA 的加工。

每个转录单位转录成单个的 RNA 前体分子，经剪切后变成为成熟的 RNA 分子。若不进行剪切，前体分子是沉降系数为 30S 的 RNA。所有的转录单位都有双重启动子(double promoter)结构。第一个启动子在 16S rRNA 起始点上约 300bp 处，可能是基本的启动子。在不同的 *rrn* 操纵子中转录单位前面的 150bp 是不同的。在此区离 P1 110bp 有第二个启动子 P2。16S 和 23S rRNA 侧面的间隔序列是保守的(图 13－4)。

在 16S 和 23S 之间是 400～500bp 的转录间隔序列(TS)。在此区域中有一个或多个 tRNA 基因。在 4 个 *rrn* 操纵子中的两个转录间隔中含有单个的 tRNA 即 $tRNA_2^{Glu}$。在其他 3 个 *rrn* 操纵子中，TS 含有 2 个 tRNAs ($tRNA_1^{Ile}$ 和 $tRNA_{1B}^{Gla}$)。

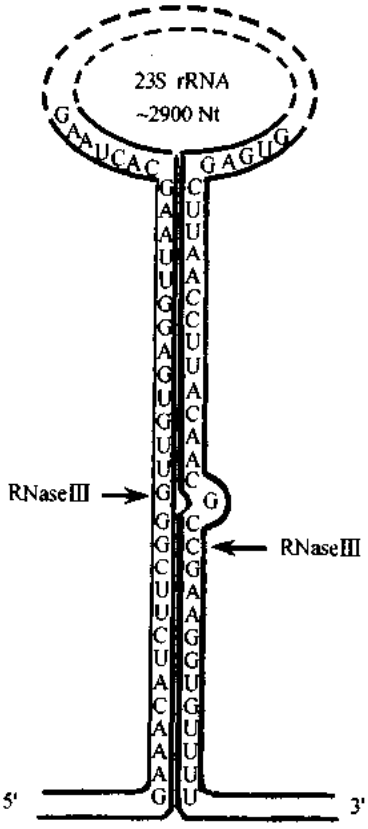


图 13－5 RNase Ⅲ切割 23S rRNA 前体产生 P23S rRNA 。(依 B. Lewin: Genes Ⅵ, 1997, Fig. 12－51)

rRNA 前体的加工是由 RNase Ⅲ负责的。在 *rnc*－的细胞中不存在成熟的 rRNA，只积聚了 30S 的前体 rRNA 分子。这种前体分子在体外能被 RNase Ⅲ切成成熟的 16S，23S 和 5S rRNA(图 13－4，表 13－1)。

已知 P^{16} 和 P^{23} 是 16S 和 23S rRNA 的前体。每种前体分子都要比核糖体中的 rRNA 分子要长得多，这是由于其 5′和 3′端的序列在成熟时还要被剪切掉。

RNase Ⅲ由 2 个亚基组成，不含 RNA，起内切酶的作用。它可能识别一级结构和二级结构相结合的某种特征。在 30S 前体 RNA 中，P16S 和 P23S 各自的 5′和 3′都能互补配对，形成含有 1600nt 和 2900nt 的茎环，RNase Ⅲ就在二者的茎上交错切割 (图13－5) (而不在单链泡上切开)产生了 P16S，P23S 和 P5S rRNA 前体。再进一步由外切酶进行修正，切除多余的部分(图 13－4，表 13－1)形成各种成熟的 rRNA 分子。

表 13－1 *E. coli* 中含有的小分子核酸酶

酶	基因	底物	功能
RNaseP	*rnpA*	tRNA 5′端	内切
RNaseO	*rnpB*		
RNaseBN	?	tRNA 3′端	外切
RNaseD	*rnd*	tRNA 3′端	外切
RNaseT	?	tRNA 3′CCA	外切
RNaseⅢ	*rnc*	rRNA 和 mRNA	内切
RNaseR	?	rRNA 和 mRNA	外切
RNaseE	*rne*	5S rRNA	内切
RNase Ⅰ	*rna*	大部分 RNA	内切
RNase Ⅱ	*rnb*	RNA	外切
多核苷酸磷酸化酶	*pnp*	RNA	外切
RNaseH	*rnh*	RNA-DNA 杂合链	内切

引自 B. Lewin: Genes Ⅳ, 1990, Table 14.2

二、真核的 tRNA 和 rRNA 的加工

1. tRNA 的加工

真核 tRNA 的基因和原核不同：(1)真核的前体分子 tRNA 是单顺反子，但成簇排列，基因间有间隔区。如在爪蟾中每个基因组中各个 tRNA 基因超过 200 拷贝，是一种重复序列；(2) 真核 tRNA 基因一般都比原核 tRNA 基因多得多，如酵母约有 400 个 tRNA 基因；(3) 5′端全有单磷酸核苷酸，表明已被加工过；(4) tRNA 的前体分子中含有内含子，其特点是：①位置相同，都在反密码子环的下游（图 13－6）；②不同 tRNA 的内含子长度和序列各异；③外显子和内含子交界处无保守序列，不符合 Chambon 法则，故其内含子的剪切是依靠 RNase 异体催化。进行剪切（见本章第四节）；④内含子和反密码子配对形成茎环，改变了反密码子臂的结构，保护了反密码子，使其免受某些酶的降解。

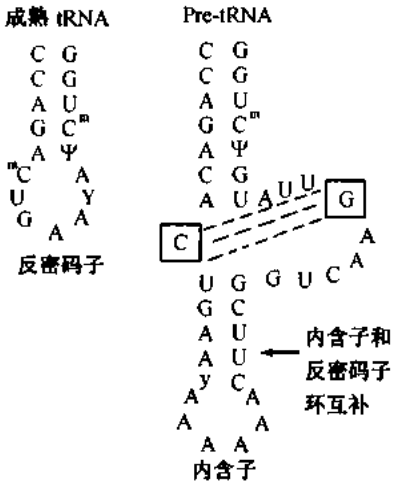


图 13－6　酵母 tRNA PHE 前体分子中内含子和反密码子配对改变了反密码子臂的结构。茎上和环上的个别碱基是剪切所需要的。(B. Lewin: Genes Ⅵ, 1997, Fig. 36－24)

真核 tRNA 的加工和原核相似，但也有区别：①真核 tRNA 前体中无二聚体和多聚体；②增加了剪接内含子的过程；③都要加 CCA。

真核 tRNA 的加工多以酵母为材料进行研究。同样通过诱变可以获得 tRNA 加工的温度敏突变型，来获得未加工的 tRNA 前体，然后在体外加入野生型酵母细胞的抽提

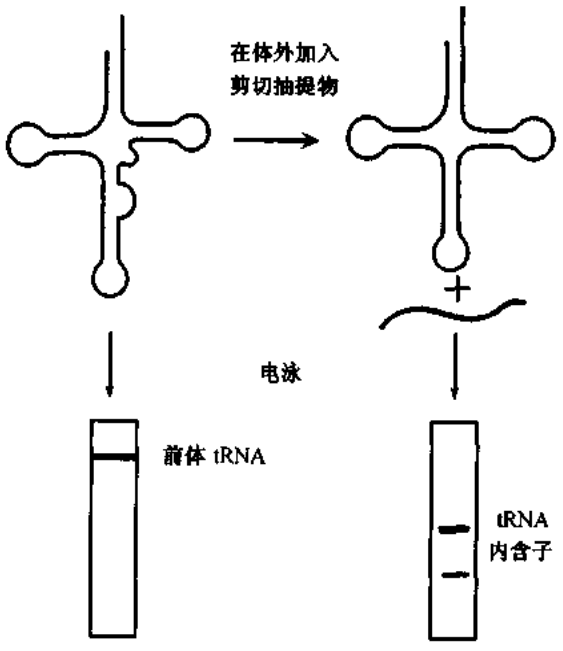


图 13－7　酵母 tRNA 前体的体外剪接。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 30－25）

物和控制 ATP 来观察，结果发现酵母 tRNA 的加工主要分成二步：①切除内含子；②连接外显子。

在未处理前先进行凝胶电泳，结果只显示一条带，且跑得较慢，显然这是 tRNA的前体分子；当加入核酸内切酶后无需加入 ATP，反应后再走电泳，结果出现二条带，一条是剪切后游离出的内含子，另一条是互补的外显子，称为 **tRNA 的半分子**(tRNA half -molecules)（图 13－7)。切除 tRNA 内含子的核酸酶很特殊，不是形成 3′－OH 和 5′－P，而是产生了 **2′－3′环磷酸和 5′－OH**，因此不能直接连接。此步反应无须 ATP(图 13－8)。接着必须通过**环磷酸二酯酶**(phosphodiesterase)将 2′－3′环磷酸打开，形成3′－OH，2′P。5′端须在激酶作用下将 5′－OH 磷酸化，再通过连接酶将两个半分子连接起来(图 13－8)。5′磷酸化需要 ATP 的参加。

哺乳动物的 RNA 连接酶可将 2′－3′环磷酸与 5′－OH 直接连接起来，无需环磷酸二酯酶和激酶的介入。

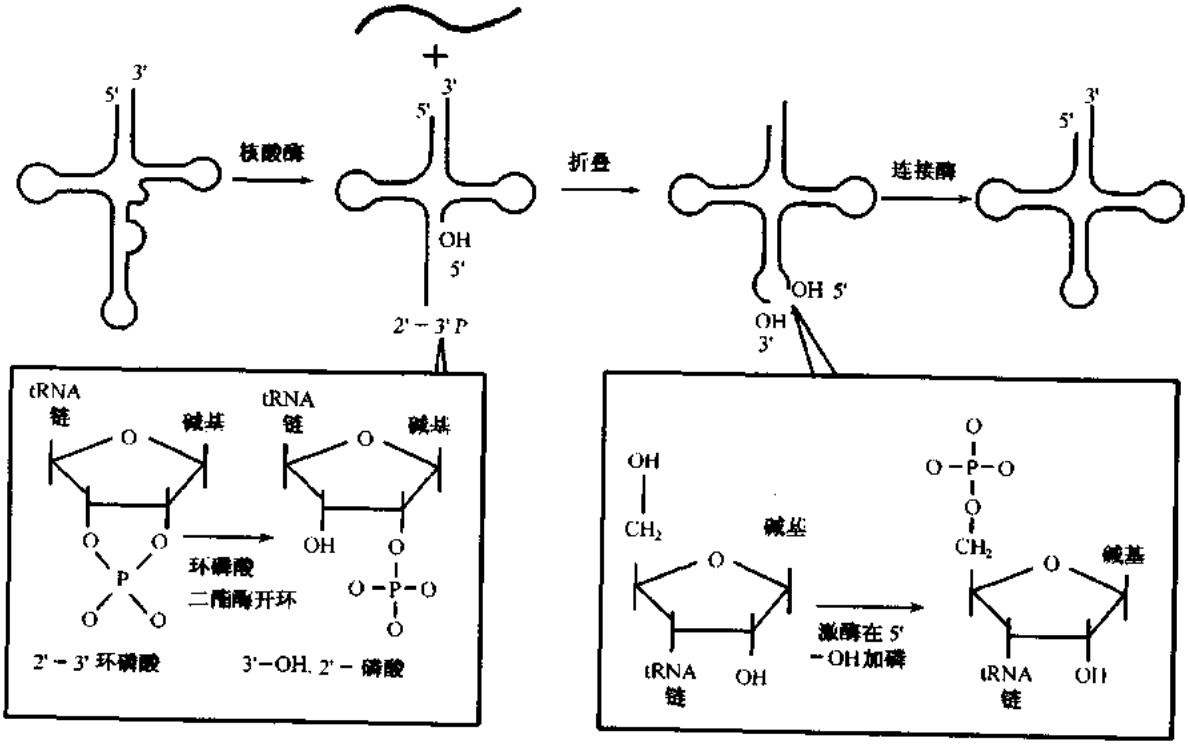


图 13－8 酵母 tRNA 内含子的剪切和连接。注：在前体分子连接前 3′端应为－OH，原图误为 P。
(B. Lewin: Genes Ⅵ, 1997, Fig. 30－26)

真核 tRNA 内含子的切除和其他内含子的切除是不同的：①即没有交界序列，也没有内部引导序列；②是依赖于蛋白性质的 RNase，而不是核糖拟酶或 snRNP；③反应的本质不是转酯反应。

真核 tRNA 内含子的精确剪切以什么为信号呢？内含子的序列和大小都不重要。内含子的突变并不影响剪切。其剪切原则上依赖于对 tRNA 共同的二级结构的识别(Johnson et al., 1980)，而不是内含子的保守序列。分子不同部分的区域对于剪切是重要的，包括受体臂，D 环，TψC 环和反密码子环。

酵母 $tRNA^{Leu}$ 基因的内含子插入一大段 *E. coli* 的乳糖操纵基因中的 21bp 后，在爪蟾胚泡(germinal vesicle)抽提液中离体转录，转录产物中即有 tRNA，也含有 21bp 外源

序列的前体 tRNA 分子。此表明内含子结构改变不影响转录和其转录后的正确剪切(Johnson et al., 1980)

前体 $tRNA^{Ile}$, $tRNA^{Leu}$, $tRNA^{Tyr}$ 及 $tRNA^{Phe}$ 等内含子的上游切点都是在反密码子 3′端之后的一个核苷酸上，另一切点常在不配对的泡上。这些 tRNA 的反密码子和内含子的序列都不同，可见剪切的核酸酶不识别一级结构，而是依赖于 tRNA 的二级结构(Peeble et al., 1983)。

切点的碱基或序列是否是剪切信号呢？现在来看 Colby 等(1981)的实验。$tRNA^{Tyr}$ 的反密码子为 5′GUA3′，当其突变为 UUA 时使可与终止密码子配对而抑制无义突变。当突变了的反密码子 3′的 A 再突变为 G 后，它的抑制效率大大下降，同时细胞中此种 tRNA 的前体大大增加(7 倍)，说明很多前体不能剪切，但剪切的位点仍是正确的(图 13-9)，实验结果表明切点的碱基只影响剪切效率，不影响酶的识别。

现在发现内含子环中的一个碱基与茎上的一个碱基进行配对对剪切来说是重要的(图 13-5)。影响此配对的其他位点的突变也会影响剪接(如产生另一种配对)。此

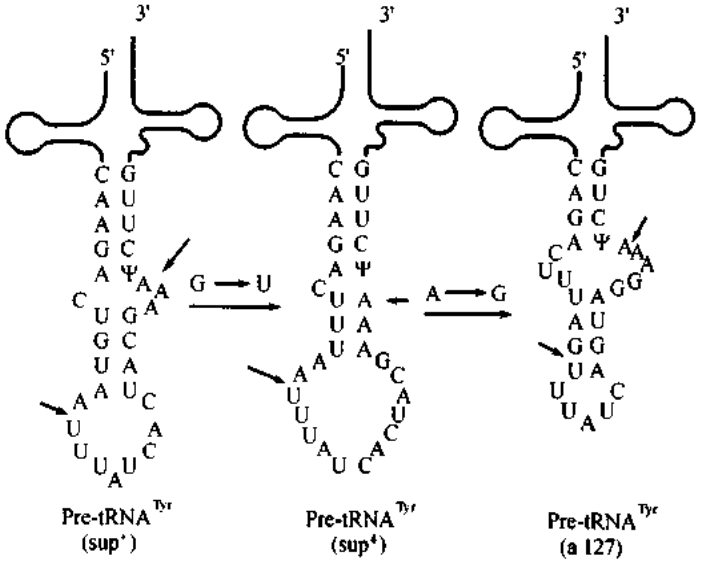


图 13-9　酵母 Pre-tRNA 反密码子和切割位点的突变。(Colby, D. *et al.*: Proc. Natl. Acad. Sci. USA., 1981, 78: 415-419)

标准有利于控制 tRNA 前体的精确剪接，这种情况类似于 tRNA 对氨基酸识别的副密码子(见第 15 章)。

2. 真核 rRNA 的加工

真核生物的 18S、5.8S 和 28S rRNA 基因是串联在一起形成一个转录本，初始转录本为 45S 前体，5S RNA 是和它们分开转录的，这和原核的 rRNA 基因不同。

在真核的 rRNA 加工中 rRNA 和其他的真核基因也不同，它没有内含子，因此加工时，无需剪切内含子这一步。真核生物 rRNA 前体的加工过程要通过 4 个阶段，所以速率相对较慢，因此主要的中间产物可从各种细胞中分离出来。从而使哺乳动物的 rRNA 加工过程得以清楚地描述。

哺乳动物的初始转录本(45S)含有 18S、5.8S 和 28S rRNA，在转录时或转录后不久约

有 110 个甲基化酶立即结合到转录本上，这些位点几乎都在核糖体上的相应位置上。

这些甲基化酶从 rRNA 加工一直保留到它成为成熟的 rRNA，其中 39 个原来的甲基酶化保留在 18S RNA 上，另外的 4 个在晚些时候在胞质中加上的。约 74 个甲基化酶结合在 28S rRNA 上，表明甲基化用于标明转录本的加工区域。

从中间产物的大小来看 rRNA 前体的加工可能有多种途径，但并不是所有的途径都已搞清楚了。图(13－10)表明 HeLa(人类)细胞和 L(鼠)细胞中 rRNA 的加工途径：①切除 5′端的前导序列，即外部的转录间隔序列(ETS)，使 45S 的初始转录本变成 41S 的中间产物；②从 41S 的中间产物中先切下 18S 的片段，但在这一步 HeLa 细胞途径和 L 途径不同，前者的切点在 18S 和 5.8S 之间的 ITS(内部转录间隔序列)，产物分别为 20S(含 18S rRNA 片段)和 32S；而后者的切点在 18S 序列和 ITS 的交界处，这样 18S rRNA 就已经成熟，无需修整，故称为**先成熟**；③部分复性：两种途径中的 32S 和 36S 中间产物(含 5.8S 和 28S rRNA)中的 5.8S 和 28S 之间进行复性，形成发夹结构；④最后修正：在 HeLa 途径通过外切酶等将 20S 中残余的 ITS 切除，还要将已复性的 32S 中的 IRS 切除掉；在 L 途径中只需将已复性的 32S 中的 ITS 切除即可。我们现在尚不知道加工的细节；不知道核酸酶被激活的信息；但已知道 45S rRNA 立即和蛋白质结合，在核糖体上进行加工，而不是以游离的 rRNA 进行加工的。

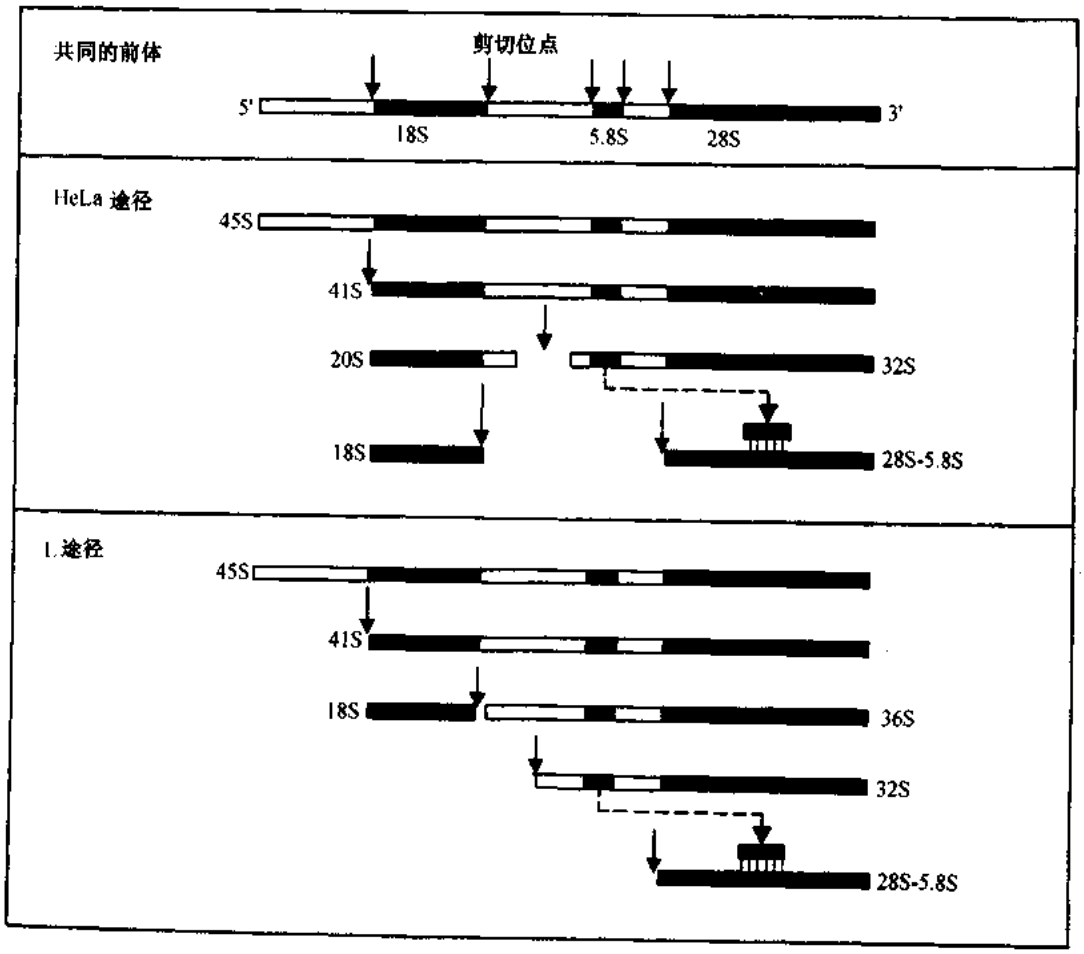


图 13－10　人类(HeLa 细胞)和小鼠(L)的 rRNA 转录后加工。(引自 B. Lewin: Genes Ⅳ, 1990, Fig. 14－13)

第二节　前体 mRNA 的加工

mRNA 前体分子的加工主要指真核 mRNA，原核的 mRNA 一般不经过加工。

真核 mRNA 的加工一般要经过四步：①5′加帽；②3′加尾；③切除内含子（内含子的切除是十分复杂的，将在下一节中进行系统的介绍）；④修饰：对某些碱基进行甲基化。

一、原核前体 mRNA 的加工

原核生物的 mRNA 很少经过加工，由于转录和翻译偶联，一般情况下一边转录一边就进行翻译，中间没有可加工的间歇时间。但在少数情况下多顺反子 mRNA 先被内切酶切成较小的单位再作为翻译的模板。例如 *E.coli* 基因组上 89′～90′处有一个操纵子含有 4 个基因：*rplj*（*L10*），*rplL*（*L7/L12*）（核糖体大亚基蛋白）和 *rpoB*（RNA 聚合酶 β 亚基），*rpoC*（RNA 聚合酶的 β′亚基）。它先转录单个的多顺反子 mRNA，然后经 RNase Ⅲ将其切开，四个基因两两分开，最后再各自进行翻译（图 13－11a）。

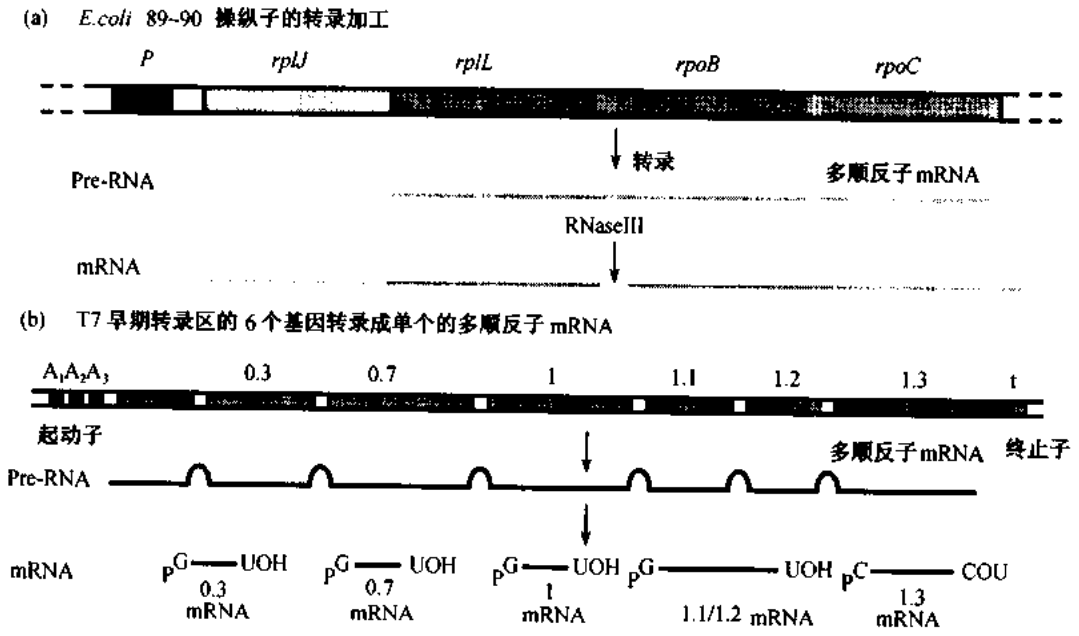


图 13－11　*E.coli* 和 T7 mRNA 的转录后加工。（a）操纵子中含 rplJ（L10）rplL（L7/L12），rpoB（RNA pol β 亚基）和 rpoC（RNA pol β′亚基）4 个基因（b）T7 早期转录区的 6 个基因产生的多顺反子 mRNA。（仿 B.Lewin：Genes Ⅵ，1997，Fig.12－46）

T7 噬菌体早期转录区中含有 6 个基因，它们同在一个转录单位中，转录后产生一个大的多顺反子 mRNA，每个 mRNA 之间在着茎环结构。由 RNase 在茎环处切开前体 RNA 分子，形成单个的 mRNA 进行翻译（图 13－11b）。RNase Ⅲ对 T7 的切割和 rRNA 前体的切割有所不同。切点在不配对的“泡”上，而不在茎上（图 13－12）。

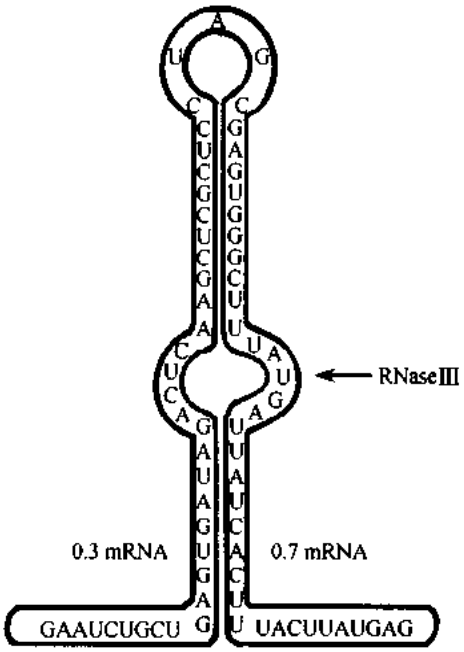


图 13－12　RNase Ⅲ在 T7 RNA 上的一个短发夹上的切割位点。(仿 B.Lewin: Genes V, 1994, Fig. 12－47)

二、真核 mRNA 前体的加工

1. 核内不均一的 RNA

在真核细胞核内可以分离到一类含量很高，分子量很大但不稳定的 RNA，称为核内不均一 RNA(heterogenous nuclear RNA，hnRNA)，其平均分子长度为 8～10kb，长度变化的范围从 2kb 左右到 14kb 左右。比 mRNA 的平均长度(1.8～2kb)要大 4～5 倍。估计 hnRNA 仅有总量的 1/2 转移到细胞质内，其余的都在核内被降解掉。人们经分析认为它是 mRNA 的前体，证据是：

(1) hnRNA 和 mRNA 有相同的序列。用小鼠核内 hnRNA 分离出的 1.5kb(15S) 的 RNA 分子和 10S β－珠蛋白 mRNA 都可与小鼠的珠蛋白基因进行分子杂交，形成 R 环。实验结果表明 hnRNA 中存在与 mRNA 相同的序列，但比 mRNA 更为复杂。此是 hnRNA 是 mRNA 前体的最有力的证据(图 13－13)。

(2) hnRNA 在体外能作为模板翻译蛋白，这也是 hnRNA 的 mRNA 前体的直接证据。

(3) 两者 5′端都有帽子结构，其他的 RNA 分子和前体都没有这种特殊的结构。

(4) 两者的合成同样不为低剂量放线菌素 D 所抑制，但能被高剂量所抑制，表明两者为相同的聚合酶所合成。

两者的 3′端都有多聚腺苷有尾巴，这也是其他的 RNA 分子所没有的。

根据以上证据表明：hnRNA 和 mRNA 之间关系密切，至少是部分 hnRNA 是 mRNA 的前体。但由于它已有 5′帽结构和 3′poly(A)尾巴，所以它不可能是初始转录本，而已经通过初步的加工。在上述的分子杂交中可以观察到 β－珠蛋白的 15S rRNA 与 DNA 杂交形成的环是完全互补的，而 10S rRNA 与 DNA 的杂交 DNA 出现了双链的环，这表

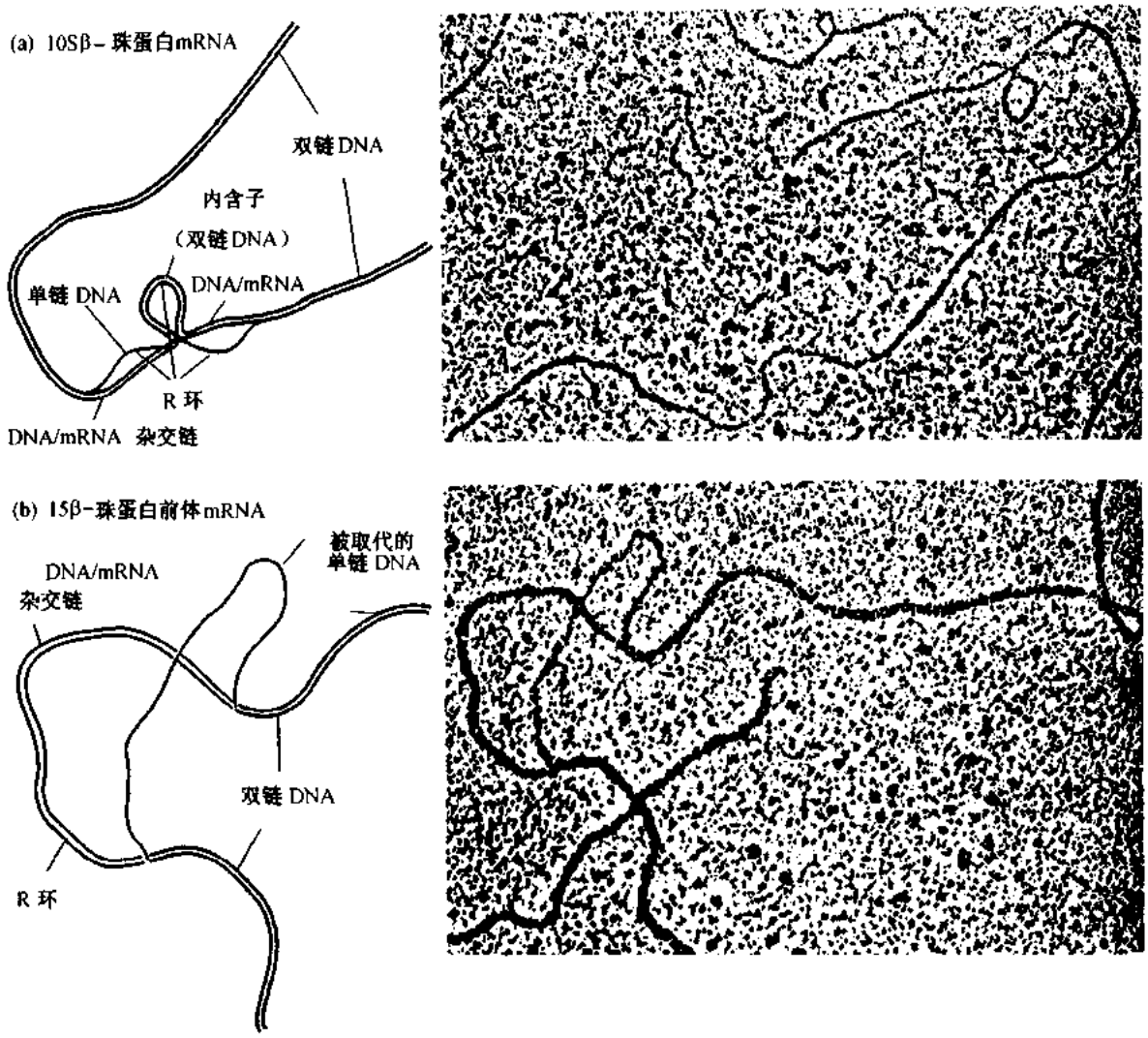


图 13-13　15S β珠蛋白前体 RNA 以及其 15S β珠蛋白 mRNA 与小鼠β珠蛋白基因杂交形成 R 环的电镜照。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 13-7）

明 15S rRNA 尚未切除内含子，和 DNA 一样，所以可完全互补，而 10S rRNA 是已切除内含子的 mRNA，它和 DNA 杂交时可使内含子区域环出。由此可见 hnRNA 虽已加帽，加尾，但尚未切除内含子。采用 Berk 和 Sharp 作图法也可证实免疫球蛋白轻链前体上含有间插序列。

hnRNA 的结构（图 13-14）有以下特点：①5′端有帽结构；②3′端有 poly(A)尾巴；③帽结构后有 3 个寡聚 U 区，每个长约 30nt；④有重复序列，位于寡聚 U 区后面；⑤有茎环结构，可能分布于编码区（非重复序列）的两侧；⑥非重复序列中有内含子区。

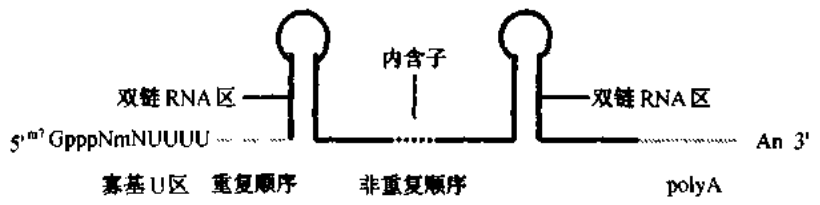


图 13-14　hnRNA 的结构。

2. 真核 mRNA 的前体加工

图 13－15 是 mRNA 前体的加工模型，适用于所有的真核 mRNA，加工过程可分为四个阶段：

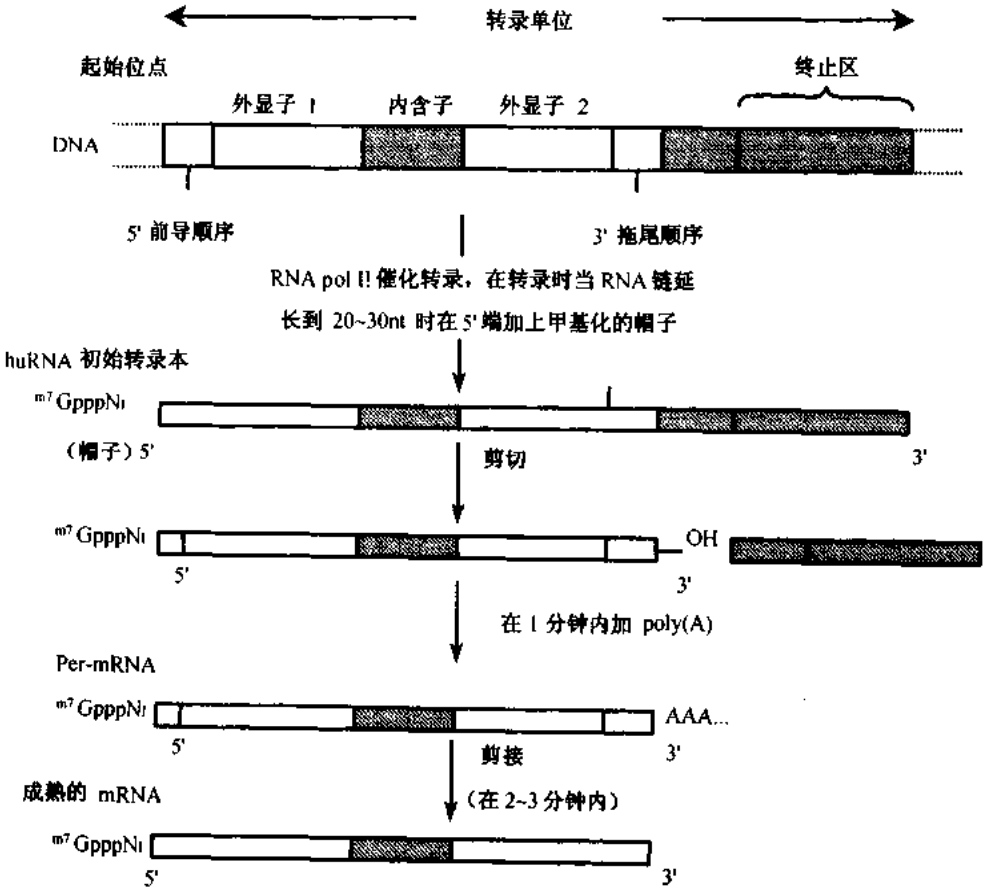


图 13－15　真核 mRNA 的加工。(仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 13－9)

(1)加帽

mRNA 的 5′端的修饰是在细胞核中进行的，所有的真核生物都是如此。转录开始是用一个三磷酸核苷（通常是一个嘌呤，A 或 G），这第一个核苷酸的 5′端仍保留 3 磷酸基因，其 3′端和下一个核苷酸的 5′位点形成磷酸二酯键，转录本的起始序列可能是：

$$5'ppp\begin{matrix}A\\G\end{matrix}\ pNpNpNp\cdots\cdots$$

但当人们在体外用酶来处理成熟的 mRNA，并未产生预期的 pppA 或 pppG，而是得到了 5′－5′三磷酸连接的 2 个核苷酸，而且末端的碱基总是一个 G，它是在转录后加在原来 RNA 分子上的。

5′末端加上鸟苷是由**鸟苷转移酶**催化的。这个反应在转录后立即发生，开始时在 RNA 的 5′端不能检测出微量的三磷酸。

反应是在 GTP 和原来的 RNA 5′端三磷酸之间进行**缩合**(condensation)**反应**。然后

再进行甲基化。此反应是在S－腺苷基甲硫氨酸的作用下对甲基化位点(如图4－16)进行加帽,如呼肠病毒,牛痘病毒。

另一类是在剪切后加帽,如疱疹病毒和口炎病毒,加帽的反应如图13－16b。

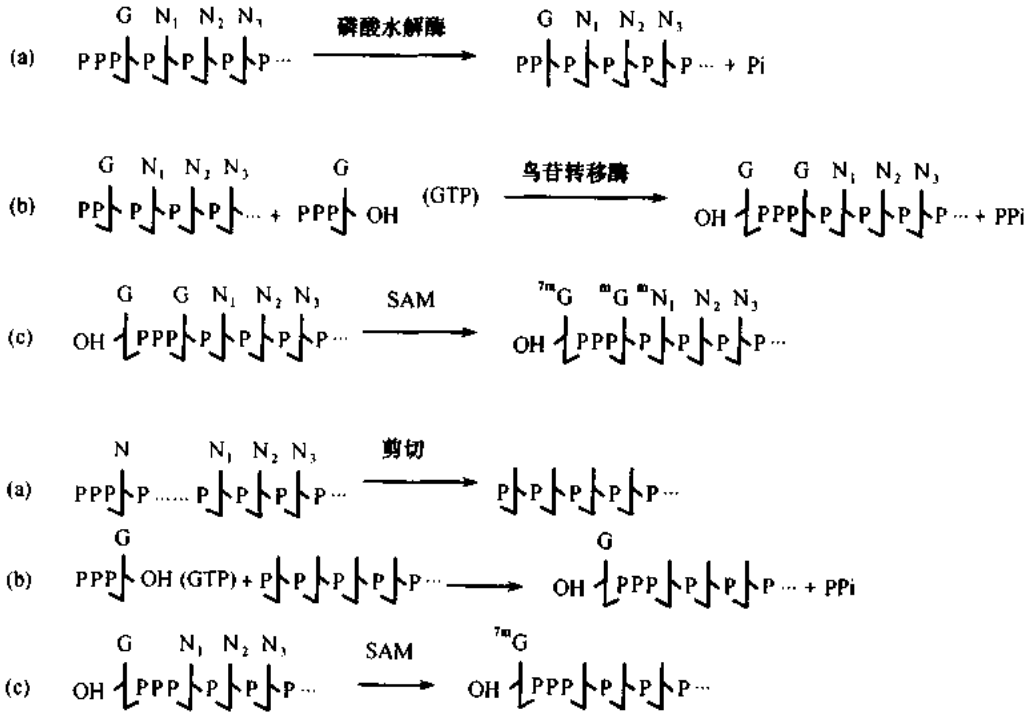


图13－16　加帽的反应步骤。(a)鸟苷转酶(加帽酶)催化在5′端加上鸟苷;(b)在S－腺苷基甲硫氨酸(SAM)的作用下进行甲基化;(c)剪切后加帽的反应。

以上在前体RNA的前导序列的5′端加上的结构就称为"**帽子**"(cap)。不同的生物中帽子结构也有几种不同的形式:在帽子上仅在末端鸟苷的第7位上存在单个甲基化位点的称**0型帽子**(cap0);除此以外在次末端核苷酸的核糖上的2′－0位点上还有一个甲基位点的称**1型帽子**;除末端以外,在第三个核苷酸的核糖上(2′－0)有甲基化位点的称**2型帽子**(cap2)(图13－17)。这三种帽子都有7－甲基鸟核苷三磷酸(m^7GPPP),它与下一个核苷酸相连的是以5′与5′方向相对的连接方式,这种特殊的结构称为**面对面核苷酸结构**(confronted nucleotide structure)。

帽子结构的功能主要有三方面:①有助于mRNA越过核膜,进入胞质;②保护5′不被酶降解;③翻译时供IF Ⅲ(起始因子)和核糖体识别(见第14章)。将有帽和无帽的呼肠病毒(Reorirus)的mRNA注入表达系统表达,结果有帽的稳定,鸟嘌呤若不带甲基时翻译效果差,但稳定性不变,但若起始密码子离5′端很远的话,或者核糖体与mRNA亲和力很强,则帽子就显得不那么重要了。

(2)*加尾*

RNA pol Ⅱ转录的mRNA前体的3′端都被切除掉,然后加上多聚腺苷的尾巴。

在高等生物中mRNA的共同特点是(酵母中无此特点)在poly(A)上游11～30nt处有一特殊序列AAUAAA,这一序列是高度保守的(图13－18)。

体外多腺苷化系统的建立和发展开辟了一条分析加尾反应的途径。3′末端结构正确的产生需要一种核酸内切酶来切割RNA,需要poly(A),还需要一种特殊组分识别

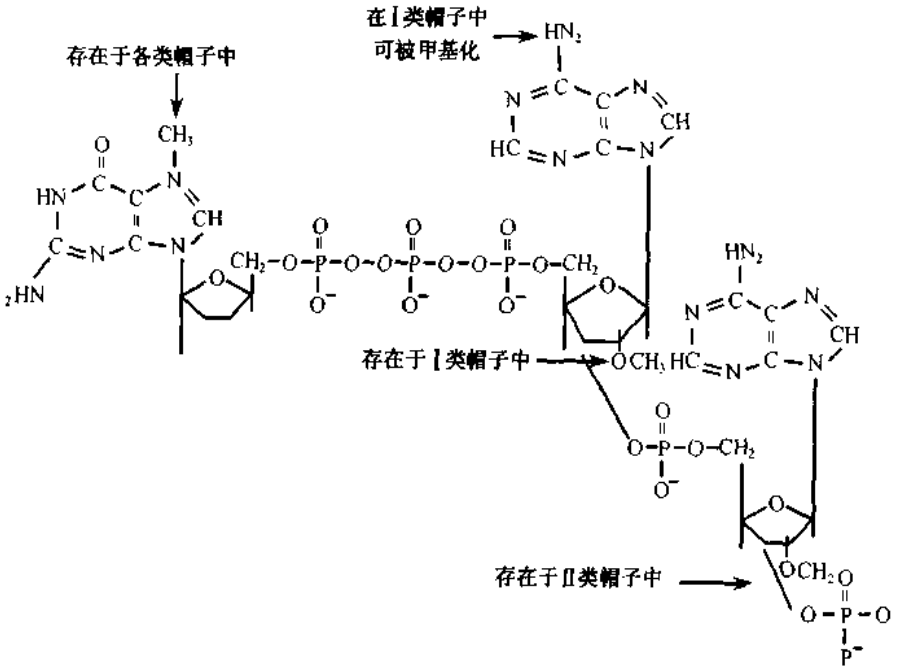


图 13－17 在帽子结构上的各个甲基位点。(引自 B.Lewin: Genes V, 1994, Fig. 10－10)

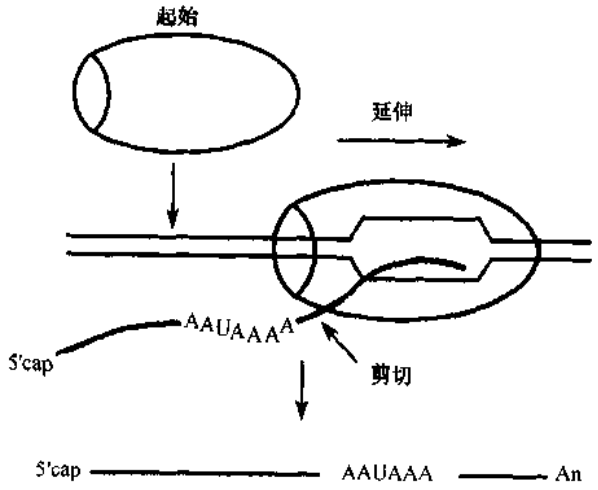


图 13－18 mRNA 3′端的剪切和加 poly A。(仿 B.Lewin:Genes V, 1994, Fig .29－14)

AAUAAA 序列并指导其他的活性。上述的内切酶尚未提纯,但特殊的组分已被阐明。

这种特殊的因子含有 3 个亚基，这 3 个亚基共同和含有 AAUAAA 的 RNA 特异结合。每一个亚基都有相同的 RNA 结合的基序，但此基序本身和 RNA 结合是非特异的。亚基之间蛋白与蛋白的相互作用可能产生了 AAUAAA 特异结合位点。

在剪切原来的 3′末端和进行多腺苷化反应中都需这种子特殊的因子。它存在于内切酶,多聚腺苷聚合酶的复合体中。这个复合体通常在特异位点切割后立即就进行**多聚腺苷化**(polyadenylation)。

poly(A)聚合酶本身提供了非特异的催化活性，只有它和别的成分组合后，合成反应才对含有 AAUAAA 序列的 RNA 具有特异性。多聚腺苷化反应要经过 2 个阶段。首先将一个短的寡聚 A 序列（10nt）加到 3′端，此反应绝对依赖于 AAUAAA 序列，这是由 poly(A)聚合酶在特殊因子指导下完成的。第二步是寡聚 A 尾巴延伸到 240nt 的长度。此反应并不需要 AAUAAA 序列，但需要一个识别寡聚 A 并指导 poly(A)聚合酶延伸的刺激因子。

有的 mRNA 是不进行多聚腺苷化的，它的 3′末端的形成和前者不同，突出的例子是编码组蛋白的 mRNA，它的 3′形成是依赖于末端的二级结构。

(3)甲基化

在真核生物和病毒的 mRNA 分子上某些位点的腺嘌呤是经过甲基化的，具体的甲基化位点还不太清楚，只知道在 HeLa 细胞，鸟类肉瘤病毒 mRNA 和 SV40 晚期转录的 mRNA 中 m^6A 主要出现在 Apm^6ApC 和 Gpm^6ApC 两种序列中，但有的真核生物和病毒的 mRNA 中并没有 m^6A，有人推测这一序列可能为 mRNA 的剪切提供信号。

第三节　内含子的剪接

内含子剪接和剪切（clearvage）是 mRNA 转录后加工中最为复杂的一环，也是近年来发展最快的研究领域之一。不同内含子的剪切和剪接，其机制，类型和酶都不完全相同（表 13－2），现分述如下。

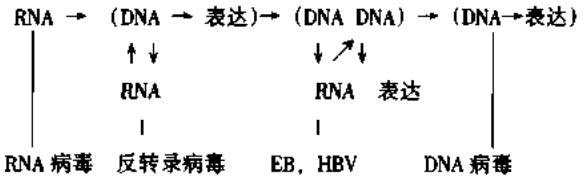
一、核酶

核酶（ribozyme）有的译为核糖拟酶，核酶等，目前尚无统一的译法，暂以核酶称之，既简单明确，也能反映其酶的本质。

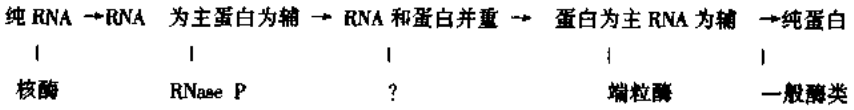
核酶是 T.Cech 和 S.Altman（1981 年）在研究四膜虫 rRNA 时发现的，1982 年 T.Cech 由核糖核酸（ribonucleic acid）和酶（enzyme）这两个词“重组”成一个新的词：“ribozyme”，它是指本质为 RNA 或以 RNA 为主含有蛋白质辅基的一类物质，这类物质具有催化功能，但和酶又有区别，主要表明在两个方面：①一般的酶是纯的蛋白质，而核酶是 RNA 或带有蛋白的 RNA；②核酶既是催化剂又是底物，随着反应的进行，本身也消失了。而酶却不同，仅催化反应，反应前后其本身的质和量都不发生改变。

核酶的发现是分子遗传学和生物化学中的一项重大突破。其意义是：①拓展了酶的概念．人们一直笃信具有蛋白质性质的酶才能催化生化反应，核酶的发现使人们了解了 RNA 本身也可具有催化功能，但和酶的催化机制不同，是一种自体催化；②揭示了内含子自我剪接的奥秘；促进了 RNA 的研究。③为生命的起源和分子进化提供了新的依据。在生命的起源和分子进化中，人们一直在争论一个问题，先有核酸还是先有蛋白？若是先有核酸，那么是什么催化其合成呢？若是先有蛋白（酶），它们合成的模板又是什么？因此形成了悖论。1971 年 Eigen 提出了一个折中的理论即核酸与蛋白质共同共化理论。核酶的发现犹如石破天惊，打破了僵局，使人们相信 RNA 可能是生命起源中最初出现的生命大分子。初期的自我复制系统可能由 RNA 单独构成。

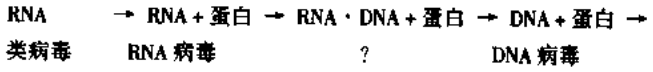
遗传信息的进化路线可能是：



酶组成的进化路线可能是：



生物体组成的进化：



核酶反映了进化的早期阶段，RNA 既是信息的载体，又具有催化的功能，在进化的过程中这两者开始歧化，将携带信息的功能交给 DNA，使得遗传信息更为稳定、丰富；将催化的功能交给蛋白，使得催化功能更为特异和多样。

核酶催化的反应分为自体催化和半自体催化两类，自体催化包括：①第一组内含子的自我剪接；②第二组内含子的自我剪切；③植物类病毒和拟病毒的自我剪切。半自体催化包括：①核 mRNA 内含子的剪切；②锥虫 SLRNA 的反式拼接；③tRNA 5′加工，此项不属于转酯反应内含子的剪接，但也是核酶的活性之一。

除核酶之外，核酸酶和由内含子本身编码的成熟酶也参与某些内含子的剪切，如真核 tRNA 内含子和酵母 *cob* 基因的内含子的 2 的剪接(图 13－19)。

酶	催化类型	加工对象	剪切/接模式	识别信号	特殊结构	反应	中间体
核酶	自体催化	植物病毒 拟病毒	剪切	13nt 的保守顺序	锤头结构和杂种 RNA	内切	
		I 类内含子	顺式拼接	分界顺序	核心顺序和内部引导顺序	转酯反应	L19 IVS
		II 类内含子	顺式拼接	分界顺序	分支顺序	转酯反应	套索
	半自体催体	核 mRNA	顺式拼接	分界顺序	分支顺序	转酯反应	套索
		锥虫 VSG	反式拼接	分界顺序		转酯反应	Y 型结构
内切酶，连接酶，核酸酶	异体催化	tRNA 5′加工	剪切 5′	发夹结构	外部引异顺序	内切	
		tRNA 内含子	剪切	茎和环上的一对碱基	内含子和反密码子配对形成的环	内切	半分子
成熟酶	自体蛋白催化	酵母 cob 内含子	剪切	?	?		

图 13－19　不同类型内含子的结构特点和剪切/接反应。

核酶的应用前景在于人们利用其自我剪切的特点人工合成各种核酶来对付有害的基因和病毒。

二、边界顺序

断裂基因发现不久，Crick(1978)就提出了一系列发人深思的问题：①剪切作用若是通过酶来进行的，那么这种剪接酶是怎样识别 RNA 上特定的位点？②剪切酶有没有特异性？对不同的 RNA 是否需要不同的酶？③切除的 RNA 是以线状还是环状存在的？切下的 RNA 的命运将如何？④内含子有无调控作用？这些问题经过 20 年来的努力基本上都有了答案。第 4 个问题前面已作介绍，现在来阐明前三个问题。首先就要介绍内含子剪切的信号。tRNA 内含子的剪切信号是依赖前体 tRNA 的二级结构和茎上与内含子环上一对配对碱基，那么其他 RNA 内含子的剪切信号是什么？首先是边界顺序。Chambon 等分析比较了大量结构基因的内含子切割位点，发现有 2 个特点：①内含子的两个末端并不存在同源或互补。这就排除了存在二级结构的可能。在剪切的初始阶段，可能直接连接；②连接点具有很短的保守序列(如图 13－20)也称为**边界顺序**。100 种内含子的 5′端都是 GT；3′端都是 AG，因此称为 GT－AG 法则(GT－AG rule)，又称为 Chambon 法则。这两个位点序列是不同的，左边的剪接位点称**供体**(donor)**位点**，右边的剪接位点称**受体**(acceptor)**位点**。这两个位点对于剪接是十分重要的，一旦发生突变无论在体内还是在体外，会抑制剪接。此法则几乎适合于所有真核生物的核基因，这意味着它们切除内含子的机制是相同的，但不适用于线粒体，叶绿体的内含子，也不适用于酵母的 tRNA 基因。

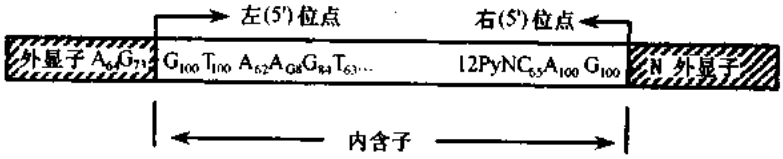


图 13－20　内含子和外显子的边界顺序——GT－AG 法则。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 30－3)

典型的哺乳动物 mRNA 有很多内含子，是什么东西保证正确的剪切和连接呢？据推测可能有两个原则：①内含子末端含有的连接位点可能是 RNA 共有的特点，这可能是特殊的序列或结构配对所需；②所有 5′位点的功能都是相同的，所有 3′位点也几乎完全相同，但剪接可能遵循这样的法则：5′位点总是和下一个 3′位点相连接。

三、Ⅰ类内含子的剪接

Cech 等 1981 年用四膜虫(*Tetrahymeno thermophila*)为材料来研究真核生物染色体的结构对基因表达的影响，他们进一步分离得到了 35S 的前体 rRNA，它含有一个长 413nt 的内含子。这个 35S rRNA 无论是否加入核的抽取物，只要加入一价或二价阳离子及 GTP 就可以在体外释放出 413nt 的线性的内含子，若继续保温，那么线形内含子又可形成环状的 RNA(图 13－21)，但加其他三磷酸核苷不行。这就意味着 35S rRNA 在 GTP 的作用下可以自我剪接。为了弄清这个问题，防止少量酶含在其中造成假象他们

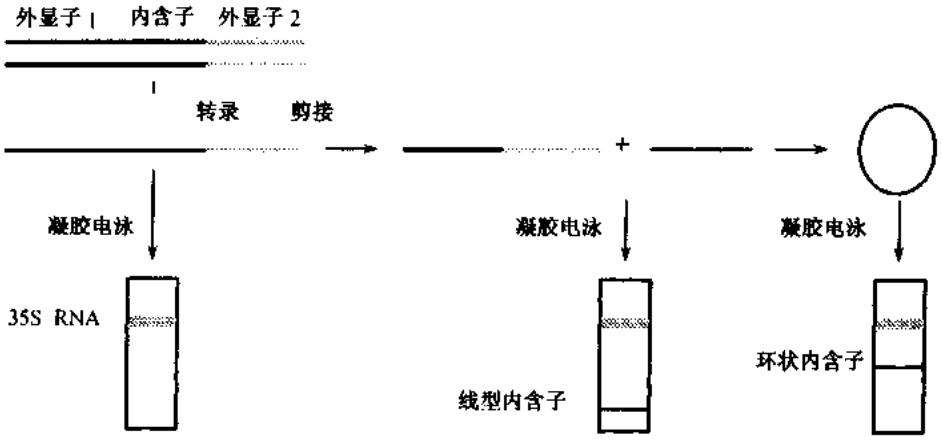


图 13－21 四膜虫 35S RNA 保温前后的电泳结果。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 31－1)

又用了大量的 SDS－酚来抽提，以较彻底地脱蛋白，或用蛋白酶来处理，但结果仍然观察到自主剪接(self-splicing or autosplicing)为了进一步证实这种自主剪接，排除其他可能的异体催化，他们用重组 DNA 技术将四膜虫 rRNA 基因中 413bp 与其两翼的序列进行分子克隆，然后再进行体外转录，用 SDS－酚来抽提，由于细菌中没有能切除真核内含子的酶类，加之 SDS－酚抽提应当说已十分雄辩地证实了四膜虫 rRNA 内含子具有自我剪切的功能。

1982 年 Cech 小组又提出了四膜虫 35S rRNA 自我剪接的模型(图 13－22)，其化学反应的本质是**转酯反应**。用同位素标记的 GTP 加入到分离的 35S rRNA 中，发现它留在 413nt 的内含子内，因此推测 GTP 上的 3′－OH 对内含子 5′和外显子交界处的 U－A 之间的磷酸二酯键，发起**亲核进攻**，切下外显子，而其 3′OH 和内含子 5′端(G)的磷酸形成了新的磷酸二酯键，从而结合到内含子上，这就是转酯反应(图 13－23)；切下的外显子，3′端的－OH 基又对内含子 3′端交界的磷酸二酯键作亲核进攻，切下了 414nt 内含子，而和外显子 2 以新的磷酸二酯键相联。切下的 414(413 1VS 加上外加的鸟苷)内含子其 3′－OH 又对本身 5′端第 15、16 两个碱基之间的磷酸二酯键进行亲核亲核进攻，切下 15nt 的小片段，余下的 399nt 的 1VS 首尾相连而呈环状；环形内含子可进一步水解，在首尾相接处切开磷酸二酯键，成为 399nt 线形内含子，因为已丢失了 15 个 nt，故称其为 L－15IVS，L 是“lose”的缩写。L－15 IVS3′端的－OH 再次对其本身 5′端第 4－5 碱基之间的磷酸二酯键发动亲核进攻，切下 4 个 nt 的小片段，余下环形的 395nt 内含子，经再次水解，产生了 395nt 线状的内含子，因前后一共丢失了 19 个 nt 的小片段，故称其为 L－19IVS。

1986 年 T. R. Cech 组内的 P. J. Grabowski 发现 L－19 可以催化五聚胞苷(5XpC)聚合为多聚胞苷(图 13－24)更加无可辩驳地证实了四膜虫 rARNA 内含子确实具有酶的催化功能，建立一种崭新的概念“核酶”。

Ⅰ类内含子常分布于低等真核生物的细胞器中，如四膜虫，绒泡菌(*Physarum polycephalum*)的 rRNA；大部分真菌如酶母的 mtDNA；植物叶绿体基因的内含子；酵母细胞色素 B 基因(*ScCOB*)，甚至在一直延伸到原核生物－T4 噬菌体的 3 个基因中也存在Ⅰ类内含子。

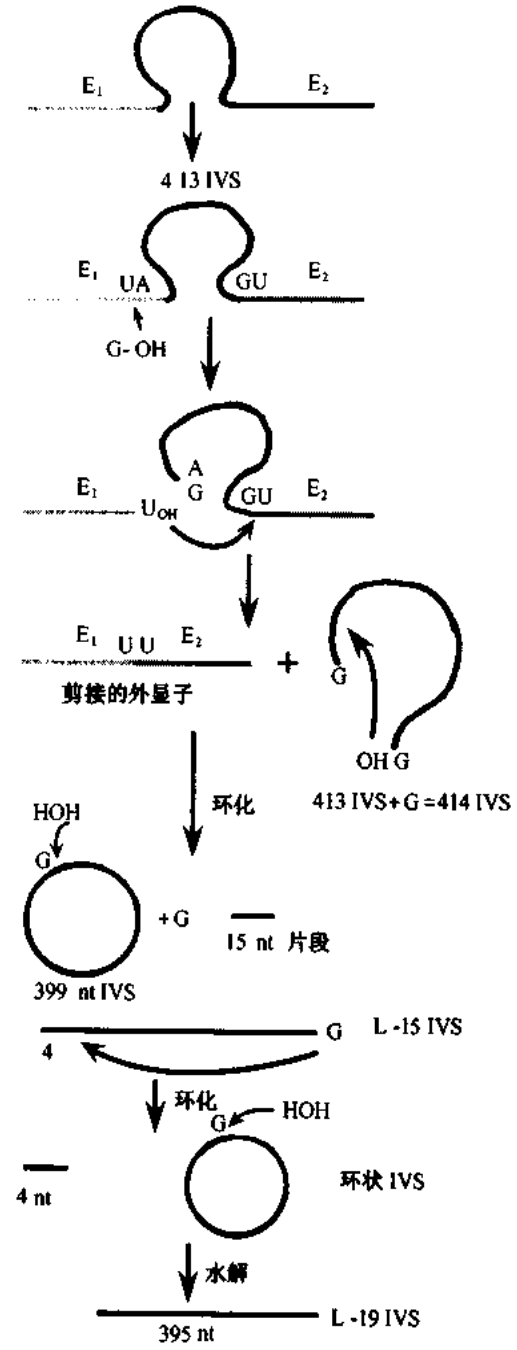


图 13-22　四膜虫 35S rRNA 以转酯反应进行自我剪接的模型。

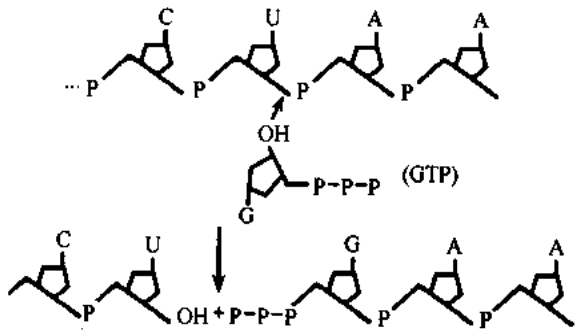


图 13-23　GTP 亲核进攻起动转酯反应。

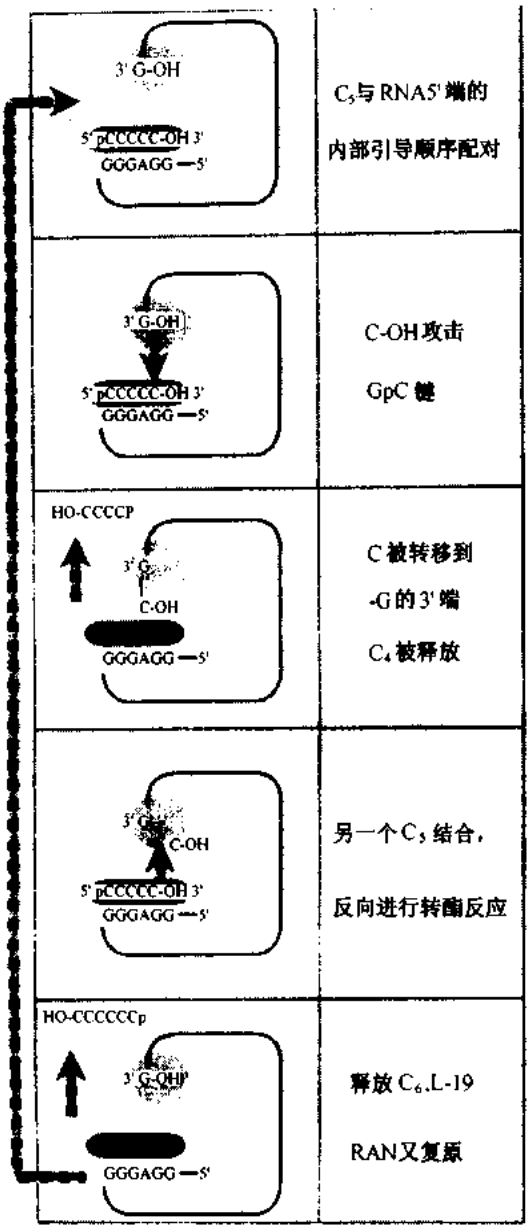


图 13－24　L－19内含子可以催化五聚胞苷(5XpC)聚合为多聚胞苷。(引自 B.Lewin: Genes Ⅵ,1997,Fig.31－8)

1. Ⅰ类内含子的结构特点

(1)其边界序列为 5′U－G 3′。

(2)内含子中具有**中部核心结构**(central core structure)。

所有的Ⅰ类内含子中都具有 2 级结构，图 13－25 表示四膜虫内含子中二级结构模型，共九个配对区(P1～P9)其中有 2 个配对区(P4 和 P7)是Ⅰ类内含子中共有的保守序列，P4 由 P 和 Q 序列形成，长 10nt，有 6～7 碱基是可以配对的。P7 是由 R 和 S 序列构成的，长 12nt，但只有 5 个碱基可配对。其他的配对区在不同的内含子中是不同的，突变分析表明，P3，P4，P6，P7 是核心结构，也就是可以执行催化的最小区域。

(3)内部引导序列 (internal guide sequence，IGS)。

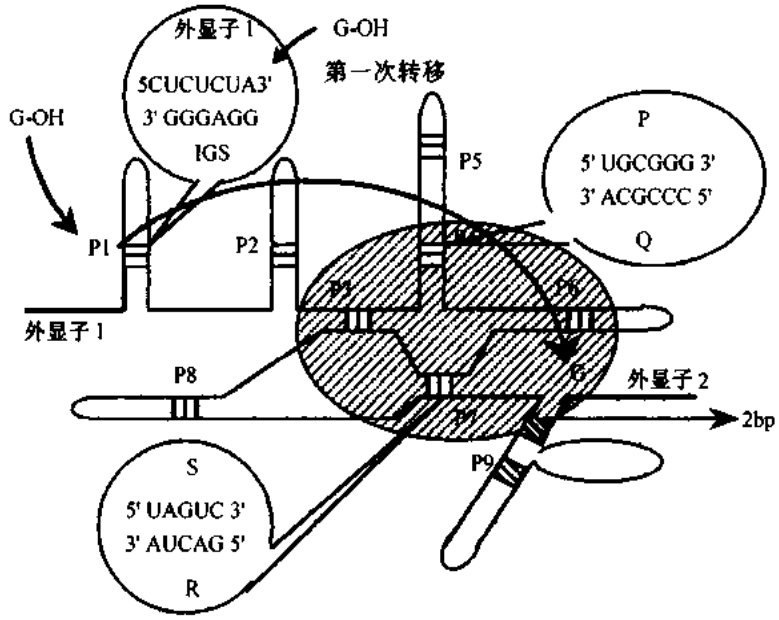


图 13－25　Ⅰ类内含子中含有的共同的二级结构。（仿 B.Lewin: Genes Ⅵ, 1997, Fig.31－4）

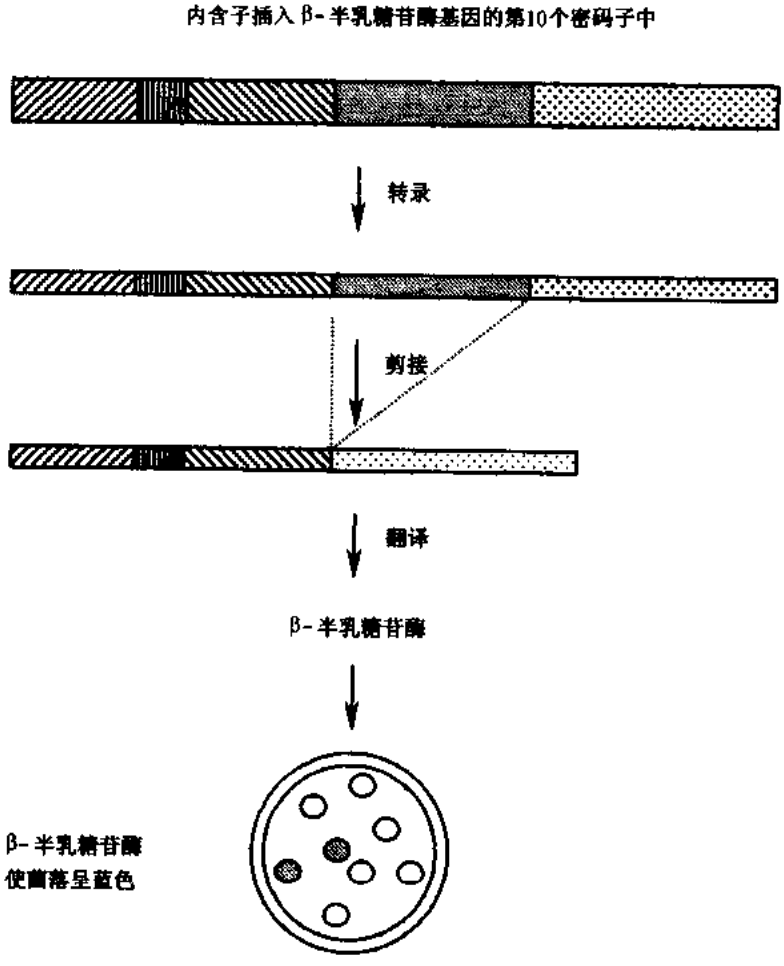


图 13－26　内含子活性的检测。（仿 B.Lewin: Genes Ⅵ, 1997, Fig.31－5）

Davies(1982年）第一个提出了内部引导序列，它由六个 nt 组成，GGAGGG(四膜虫)。

有的配对反应是直接涉及到内含子的剪切，外显子连接的酶促反应。P1 含有左边外显子的 3′端，内含子中可与外显子配对的序列称为内部引导序列。最初人们认为 IGS 的作用是通过和两个外显子近侧区域配对，使外显子并在一起，现在认为其作用是决定剪接的专一性。而且使切点的 U 处于易于受到攻击的暴露点。有的序列很短，在 P7 和 P9 之间配对，要在内含子 3′端 G 激活前立即配对。

碱基的配对对于产生核心结构是很必要的，顺式作用位点的突变就会阻止 I 类内含子的剪接。各种突变的线粒体内含子在体内都不能被切除，而四膜虫的内含子插入到细菌基因中，再转化到细菌中，它仍可以剪切。

图 13－26 表示构建一个重组 DNA，转化到 *E. coli* 表达。将自我剪接的内含子插到 β－半乳糖苷酶的第 10 个密码中再转化 β-gal⁻ *E. coli*，若此内含子不被剪接的话，β－半乳糖苷酶基因受到破坏不能表达，菌株为白色，若能自我剪切，则 β－半乳糖苷酶基因可以翻译成完整的酶，能使底物 X-gal 发酵变兰。用此方法可以检测内含子是否具有自我剪接的活性。

2. Ⅰ类内含子的剪接机制

核酶具有多种催化活性，对于Ⅰ类内含子来说这些活性是由于特殊的二级和三级结构的产生形成活性位点，相当于传统酶的激活位点。图 13－27 就表示在这些位点上进行的剪接反应。内含子的二级三级结构形成了 G 苷酸结合位点和底物结合位点，后者依赖于内部引导顺序的配对来确定。核心结构和内部引导顺序又使这两个位点彼此靠近，便于相互作用。

首先是 GTP 进入 G 结合位点，左边外显子的 3′端通过和引导顺序的互补配对进入底物位点。GTP 的 3′－OH 作用左边外显子和内含子交界处的磷酸二酯键，本身结合到内含子的 5′末端，被切下的 5′外显子仍保持在底物位点上，并未游离。第二步内含子的 G414 （即 3′交界序列上的 G）又进入了 G－结合位点，切下的外显子的 3′－OH

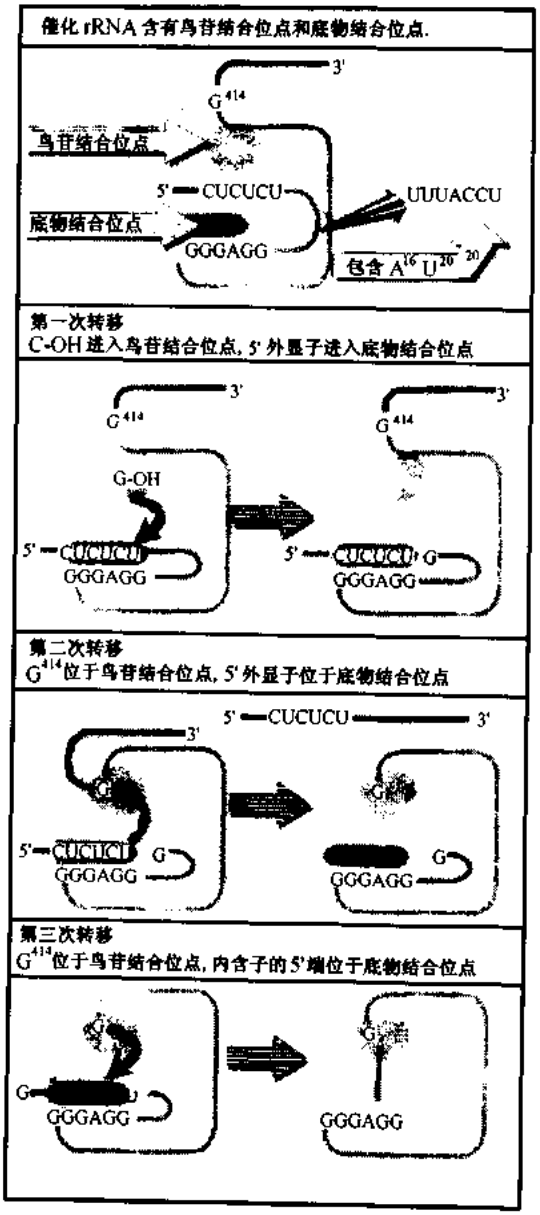


图 13－27 Ⅰ类内含子的二级结构与内含子的切除。四膜虫 rRNA 的二级结构含有鸟苷结合位点和底物结合位点。（引自 B. Lewin: Genes Ⅵ, 1997, Fig. 31－6）

又对 G－结合位点上的 G 与 3′外显子之间的磷酸二酯键发动亲核进攻，切开磷酸二酯键，而其 3′－OH 和 3′外显子的 P 重新形成磷酸二酯键而连接起来，这样就完成内含子的剪接（图 13－38）。内含子成为线形的 413nt 的 RNA 分子。第三步，内含子 5′端和引导顺序相邻的序列通过引导顺序互补配对，进入底物位点。仍然留在 G 结合位点上的 G414 其 3′－OH 再作用底物位点上的序列，切下 19nt 的内含子 5′端，并和余下内含子的 5′－P 形成二酯键，形成 414 个 nt 的环状内含子。

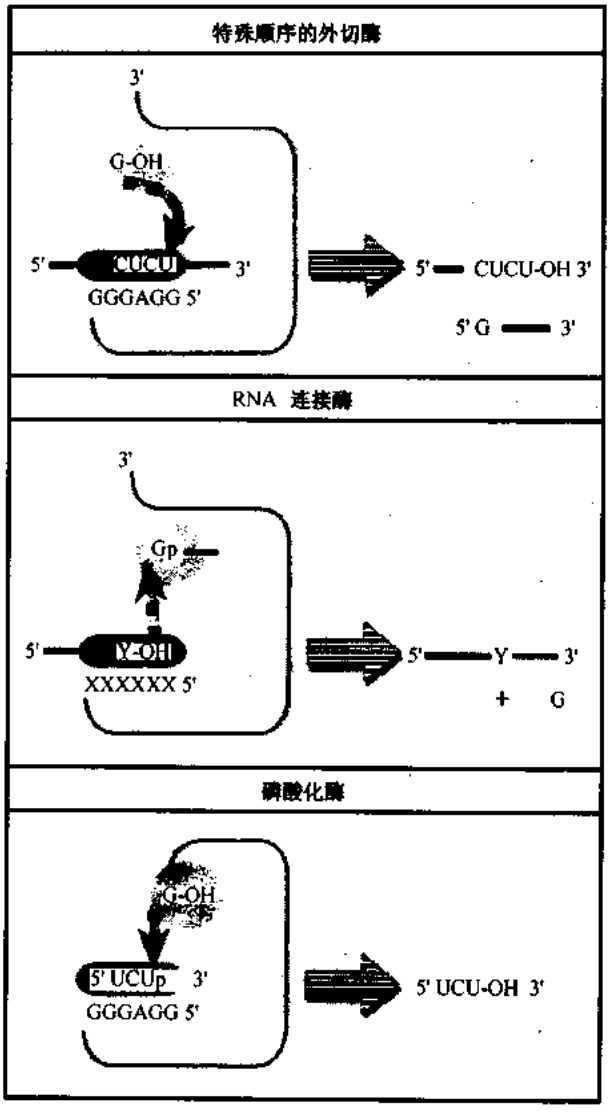


图 13－28　核酶催化内含子的剪接反应涉及到转酯反应。核酶具有多种酶活性。
（引自 B.Lewin: Genes Ⅵ, 1997, Fig. 31－9）

四、Ⅱ类内含子的剪接

Ⅱ类内含子的剪接和Ⅰ类内含子不同，而和核 mRNA 内含子的剪切有些相似。

Ⅱ类内含子分布在酵母 mtRNA，细胞色素氧化酶的 a 亚基，b 亚基，玉米 mtRNA，tRNA 以及真核 mRNA 前体中。

1. 结构特点(图 13－29)

(1) 边界序列为 5′↓GUGCG…YnAG↓,符合 GT－AG 法则。

(2) 二级结构。二级结构的形成使两个并列的功能区靠近。功能区 5 和功能区 6 被 2 碱基隔离开。功能区 6 含有 1 个不配对 A 残基,其上带有 2－OH 首先进行转酯反应。

(3) **分支点顺序** (branch-point seguence) 在内含子的 3′,也就是在第 6 功能区有一 6～12nt 保守序列,在哺乳动物中为 YNCURAY(Y:嘧啶,R:代表嘌呤,N 表示任何核苷酸)。分支点顺序可和上游序列互补,形成茎环,但 A 不配对(图 13－29)。

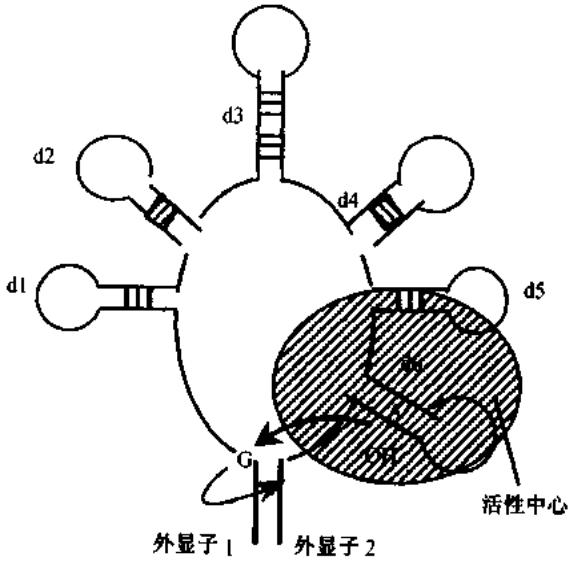


图 13－29 Ⅱ类内含子的二级结构。(仿 B.Lewin: Genes Ⅵ,1997,Fig.30－15)

2. 剪接机制

Ⅱ类内含子的剪接无需鸟苷的辅助,但需镁离子的存在。首先是分枝点 A 的 2′－OH 对 5′端外显子和内含子交界处的磷酸二酯键发动亲核进攻(图 13－30, 31)切下外显子 1,内含子 5′的边界序列上的 G 以 5′磷酸和分枝点 A 的 2′－OH 形成磷酸二酯键,从而产生了套索(lariat)结构;第二步是切下的外显子 1 其 3′－OH 继续对内含子 3′端

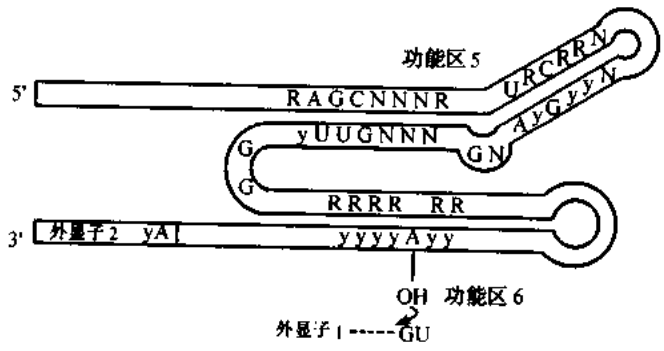


图 13－30 Ⅱ类内含子的剪接反应的活性中心。
(仿 B.Lewin: Genes Ⅴ,1994,Fig.30－17)

的交界序列进行亲核进攻，切下的外显子 2′的 5′磷酸和外显子 1 的 3′— OH 形成磷酸二酯键，连接在一起，同时释放出套索状的内含子。

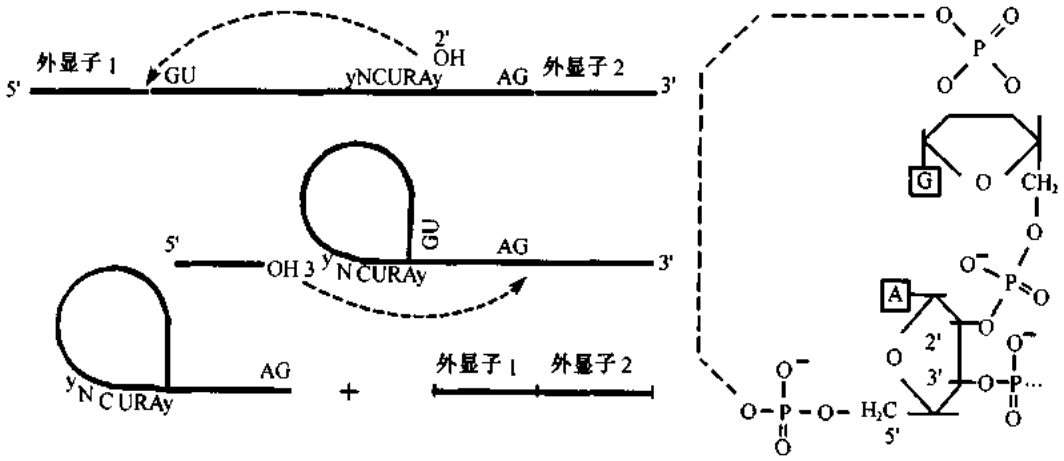


图 13－31　Ⅱ类内含子剪接机制。(参考 P. J. Bussell: Genetics, 3rd ed., 1992, Fig. 13－10)

五、核 mRNA 的剪接

核 mRNA 剪切过程(图 13－32, 33)和Ⅱ类内含子十分相似，但有一个根本的不同就是核 mRNA 前体的剪接本身不能形成二级结构，必需依赖于 snRNP(U1, U2, U4, U5, U6)的帮助才能形成剪接体，进行自我剪接。它们自己本身不能像Ⅰ类及Ⅱ类内含子那样依赖核心结构，形成茎环，将 5′和 3′剪切点拉在一起。和 snRNA 的结合是相当复杂的。但剪接反应的第一步 5′位点的剪接是由 U6 催化的，3′位点是由 5′外显子 1 的 3′－OH 对内含子 3′端切点进行转酯反应来完成。

1. 结构特点

(1)边界顺序：其边界序列完全符合 GU－AG 法则。

(2) 分枝点顺序：它和Ⅱ类内含子相似也具有分枝点顺序。位于内含子 3′端上游 18～40nt 处，序列为 $Py_{80}NPy_{87}Pu_{75}APy_{95}$ 其中 A 为百分之百的保守，且具有 2′－OH。

(3) 内含子 5′端有一保守序列 (5′GUAAGUA3′) 可以和 U1snRNA 的 5′端的保守顺序 3′CAUUUCAU5′互补。

2. 剪接机制

图 13－32, 33 说明了核 mRNA 剪接的大致过程，实际上要比这复杂得多，首先要解决的是如何识别剪接位点，一般是通过两种途径：①存在一种酶可以特异识别它们，并将两者的保守序列拉到一起；②通过 RNA 的碱基配对产生二级结构，再由一种酶来识别。

核 mRNA 的剪接可能要依赖于**反式作用 RNA**(trans-acting RNA)。在高等生物中都含有很多的不连接的小分子 RNA 片段，在高等生物中其长度在 100～300nt，在酵母中

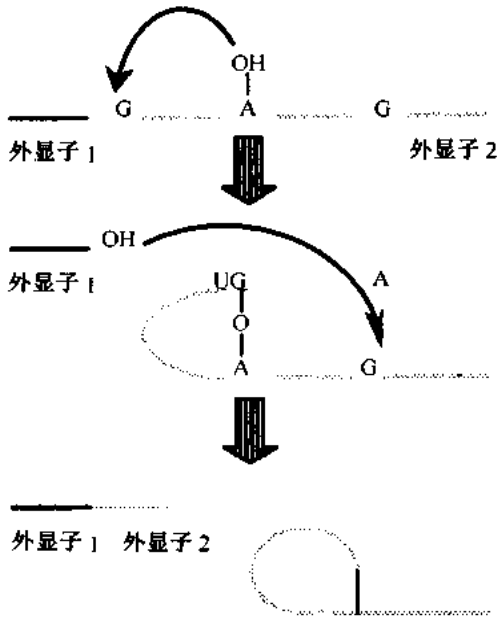


图 13－32　核 mRNA 的剪接反应。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 30－5)

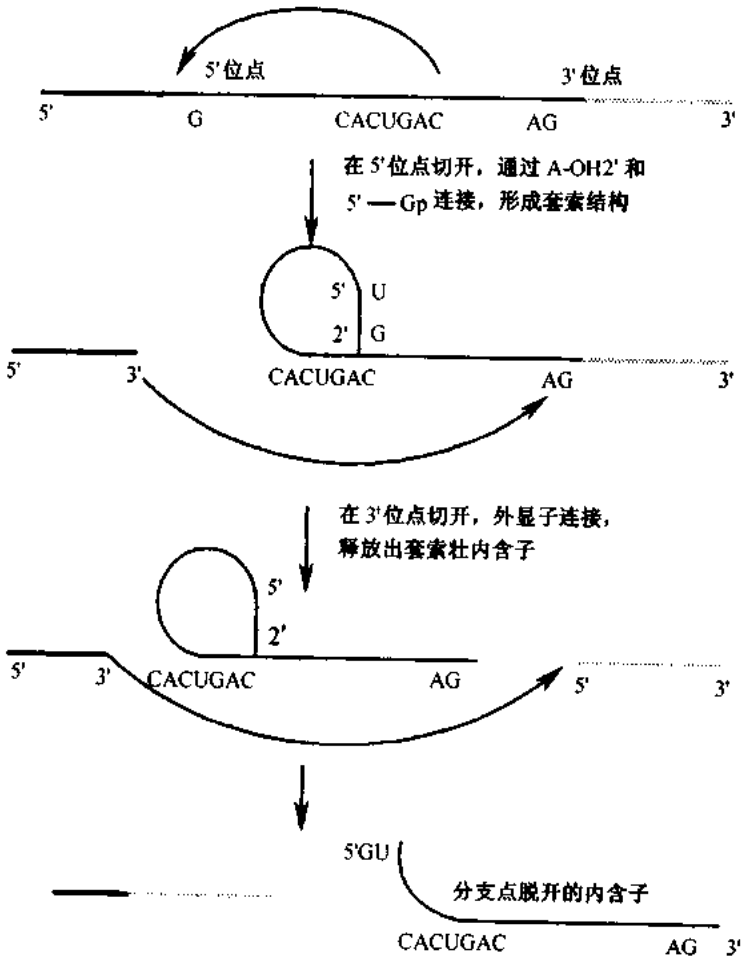


图 13－33　核 mRNA 剪接可能涉及的三个步骤。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 30－6)

长度可达 1000nt，其丰度很高。仅限制在核内的这些 RNA 称为**小分子核内 RNA**(small nuclear RNAs，snRNA)。在细胞质内的称为**小分子胞质内 RNA**(small cytoplasmic RNAs，scRNA)。在自然状态下，它们以核糖核蛋白(RNP)的形式存在。在剪接装置中会有几种 snRNPs 和大量的附加蛋白，常称为**剪接因子**，这些 snRNPs 是 U1，U2，U5 和 U4/U6。U1，U2 和 U5 snRNA 每种都含有单个的 snRNA 和几种蛋白，而 U4/U6 含有 U4 和 U6 snRNAs 及几种蛋白。U1 snRNA 自我配对形成了多个茎环结构，从而构成了不同的功能区(图 13－34)。**Sm 结合位点**是和其他 snRNP 相互作用的区域。其 5′末端有一保守序列：

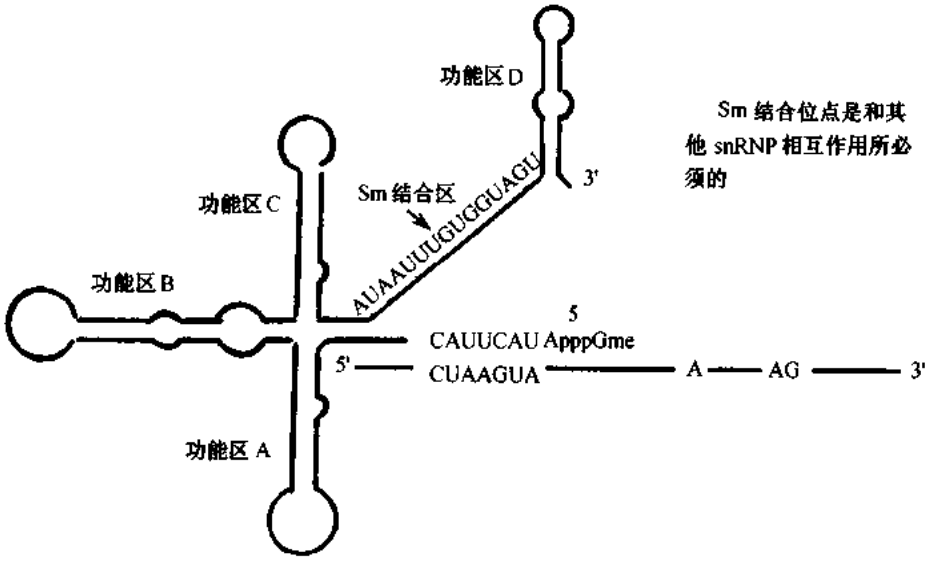


图 13－34　U snRNA 的二级结构及其功能区。

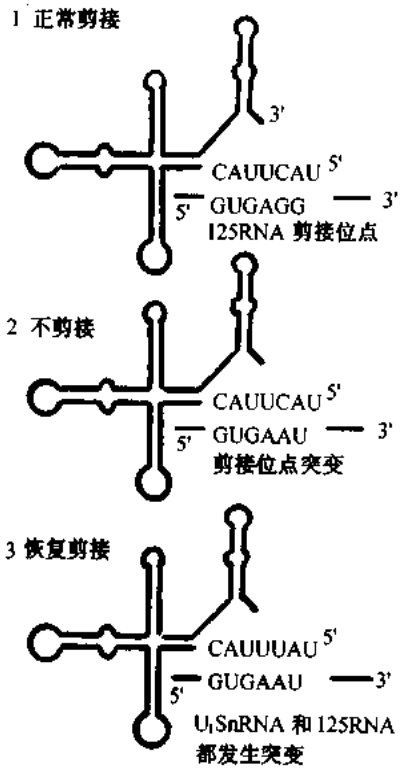


图 13－35　腺病毒 12S RNA 5′剪接位点的点变及对剪接的影响。(仿 B.Lewin: Genes Ⅵ，1997，Fig.30－9)

3′CAUUCAU 5′，这一序列可与核 mRNA 内含子 5′端的边界序列互补结合，这对于剪接反应是很重要的。以腺病毒 12S RNA 的 5′端边界顺序为例，如果此序列发生突变，从 GUGAGG 突变成为 GUGAAU，虽然仅二个碱基发生突变，配对的碱基数仍然未变，仅减少了一个氢键，结果不能剪接。若在此基础上 U1 RNA 的 5′端序列也发生相应的突变，使得突变的碱基也能配对，结果又恢复了剪接的能力。由此可见 U1 RNA 5′端保守序列和内含子 3′剪接位点的配对对于剪接反应来说是十分重要的。其作用可能是通过配对使 URNP 可以和核 mRNA3′剪接位点牢固地结合，再通过 U1 和其他 snRNP 的相互作用使 5′剪接接点与 3′剪接位彼此靠近，弥补了核 mRNA 本身不能通过形成特殊的二级结构将两个剪接位点拉近的问题。使剪接能顺利进行(图 13－35)。

图 13－36 表示 snRNP 的剪接功能和反应步骤。第一步是 U1 snRNA 通过和 5′剪接位点的配对和 5′位点结合。但 U1 也可和分枝点结合，因为带有突变的分枝位点的 pri-mRNA 其 5′位点是不能和 U1 snRNA 结合的，现在还不知道 U1 如何识别分枝位点的，可能和 snRNP 的某些组分有关，或者存在某些辅助因子。

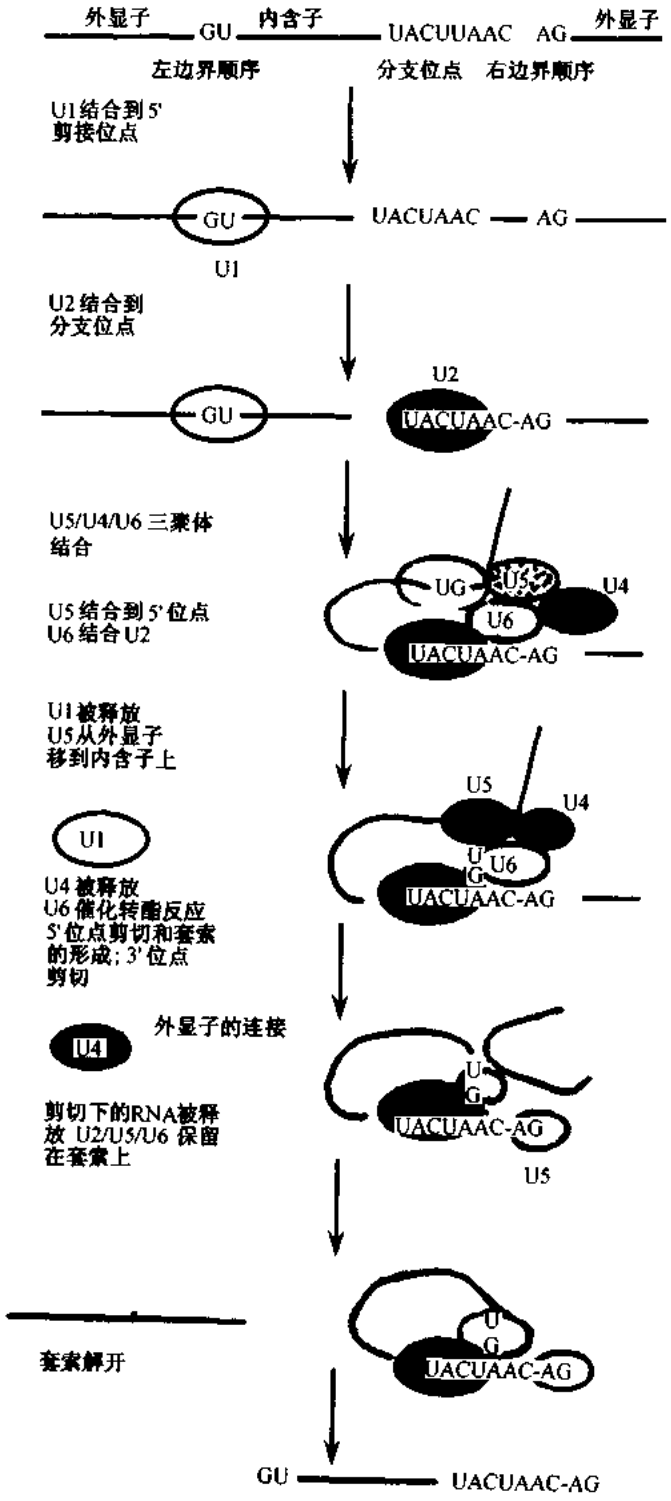


图 13－36　核 mRNA 剪接反应的步骤。(参考 B. Lewin: Genes Ⅵ, 1997。根据内容，当第二次转酯反应时 U5 改画在 3′位点上)

U1 的存在对于第二步也是必要的。在第二步中 U2 snRNA 结合在分枝位点上。U2 snRNA 通过和分枝位点的碱基配对来识别它。在酵母中分枝位点顺序为 5′UACNACA 3′,可和 U2 形成 7bp 的双链。剪接反应需要特殊的剪接因子 U2AF(U2 auxiliary factor) U2 snRNA 的结合需要 ATP 水解,并将前体 mRNA 置于剪切途径中。

由于分枝位点紧靠 3′剪接位点,U1 snRNA(结合在 5′位点)和 U2snRNA(结合在分枝位点)之间的反应使两个剪接点彼此靠近。

含有 U5 和 U4/U6 snRNA 的三聚体结合,水解 ATP,现在剪接装置的所有成分都全部到位,在剪接反应前,它们都装配起来。

U1 和 U2 是剪接起始阶段是所需要的,它们的失活会阻碍 5′位点的剪切和套索的形成。虽然 5′位点和 U1 snRNA 的碱基配对负责剪接位点的识别,但并不控制剪切。U1 在反应开始前解离。在此点上的 U5 snRNA 会改变自己的位置,开始它靠近外显子,后来移到附近的内含子上。

催化反应是通过 U4 的释放来启动,需要 ATP 的水解,U4 snRNA 的作用可能是抓住 U6 snRNA,直到 U6 参加反应。图 13-37 表明在剪接过程中 snRNA 之间的相互配对。在 U6/U4 snRNP 中,U6 有 26 个连续的碱基和 U4 两个分开的区域配对。当 U4 解离时 U6 的这个区域游离出来,产生另一种结构,其第一部分序列和 U2 配对,第二部分的序列形成了发夹结构。U4 和 U6 之间的相互作用与 U2 和 U6 之间的相互作用是互不相容的。U4 的释放启动了剪接体(spliceoome)的产生。

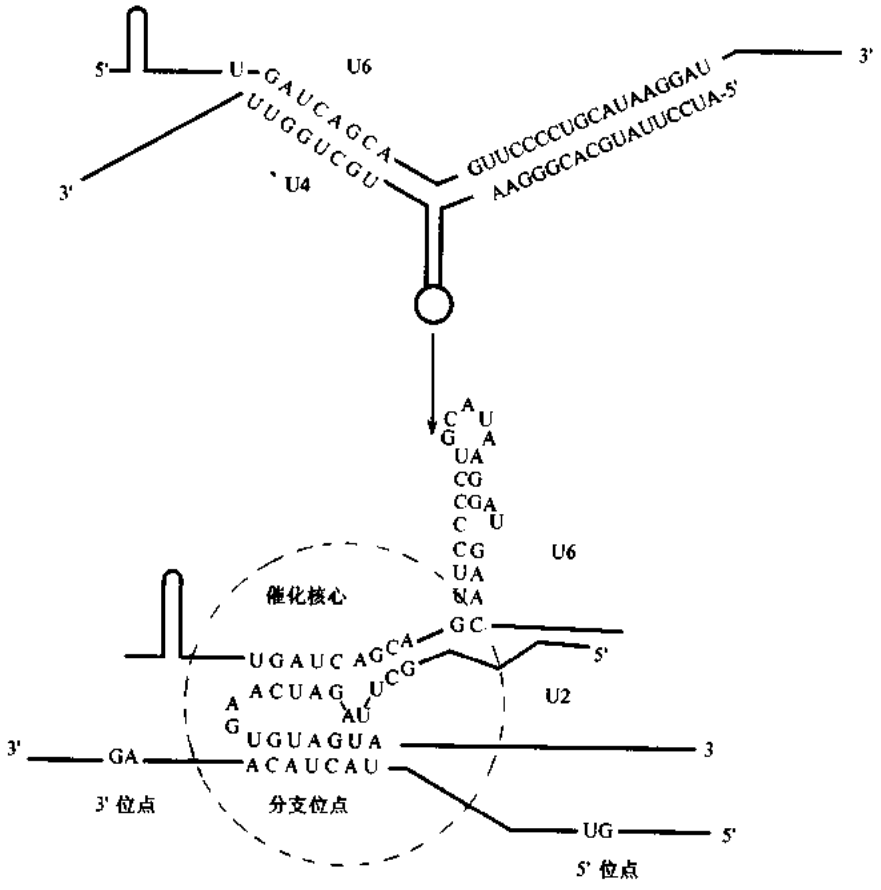


图 13-37　酵母中 U6 和分支位点的结合机制。(B. Lewin: Genes V, 1994, Fig. 31-9)

承担转酯作用的成分尚未鉴定出来，就这点而言我们还不知道剪接体的 RNA 和蛋白起什么作用。U2 和 U6 的碱基配对产生了二级结构，类似于Ⅱ类内含子的活性中心，表明催化的成分是由 U2－U6 相互作用而产生的一种 RNA 结构。与Ⅱ类内含子不同的是核 mRNA 的催化中心由 U6－U2 的介入形成，而二类内含子是由本身的第 5、6 功能区的二级结构所产生的。

分枝位点套索的形成决定了 3′剪接位点的使用，由于 3′端的保守顺序离分枝点 3′端很近，因此成为第二次转酯反应的靶子。第一次的剪切反应后很快就发生了第二次剪切反应。U5 snRNA 结合到 3′剪接位点上对于第二次剪切是必需的。但尚不清楚 snRNA 是否通过碱基配对来识别保守顺序的。

以上几种内含子的剪接机制是不同的，现总结如表 13－2 所示。

表 13－2　几种不同内含子剪接反应的区别

	酵母 tRNA	Ⅰ类内含子	Ⅱ类内含子	核 mRNA 前体
边界顺序	无	5′U↓－G↓3′	5′↓GU－AG↓3′	5′↓GU－AG↓3′
特殊顺序	C(茎上)－G(环上)	内部引导顺序	分支点顺序	分支点顺序，5′外显子顺序
二级结构	茎环构象	核心结构	5、6 功能区	连接体
基因外的成分	内切酶，连接酶	GTP，镁离子	镁离子	U1U2，U4，U5，U6
能量要求	ATP	不	不	ATP
中间型分子	半分子 tRNA	环状 L－19 IVS	套索	套索

六、反式剪接反应

内含子的剪接一般都是发生在同一个基因内，切除内含子，相邻的外显子彼此连接，称为**顺式剪接**（*cis*-splicing），但也有另一种情况，即不同基因的外显子剪接后相互连接，此称为**反式剪接**（*trans*-splicing）。

反式剪接的情况较为稀少，较典型的例子是锥虫**表面糖蛋白基因** VSG（variable surface glycoprotein），线虫的**肌动蛋白基因**（actin genes）和衣藻（*chlamydomonas*）叶绿体 DNA 中含有的 *psa* 基因。

Von der Ploeg 等（1982 年）发现在锥虫（*Trypanosome*）中许多 mRNA 的 5′端都有共同的 35b 前导序列，但在每个转录单位上游并未编码这个前导序列，1986 年 Murphy W J 和 Sutton RE 证实此是一种反式拼接。这种前导序列来源于基因组中位于别处的重复序列（200 个）所转录的小片段 RNA，每个重复单位长 135nt，前面 35nt 的前导序列可以通过转酯反应加到 VSG mRNA 的 5′端。反应如图 13－38 所示。其反应机制和核 mRNA 内含子的剪接是相似的。在 VSG mRNA 右侧内含子上有分枝点位点，可以和重复序列的内含子相连接。因为这两段序列是反式结构，所以形成 Y 型中间体而不形成套环，用去分枝酶处理就可以得到两个片段，而若是套环的话就会得到一条片段。切下的 5′外显子，即 35nt 的前导序列的 3′端可以再进行第二次转酸反应。

在线虫的肌动蛋白基因中也存在类似的情况。三种肌动蛋白的 mRNA（和某些其他的 RNAs）在 5′端有相同的 22nt 的前导序列。这个前导序列并不是由肌动蛋白基因编码，而是由长 100b 单独转录的 RNA 的一个部分。

反式剪接的 5′端外显子称为**剪接前导 RNA**（spliced leader RNA，SL RNA）。SL

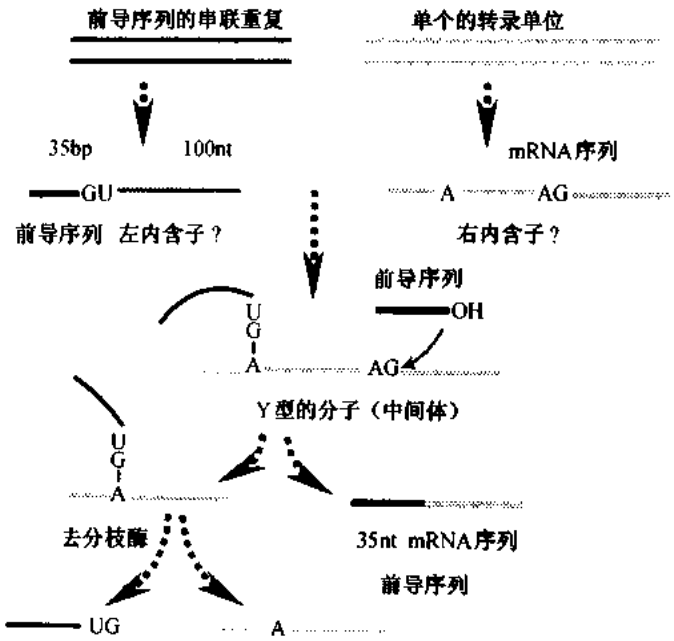


图 13－38　锥虫反式拼接的过程。（参考 B.Lewin: Genes Ⅵ, 1997, Fig. 30－22）

RNAs 存在于几种锥虫和线虫（*C. elagan*）中，它们具有共同的特点，即折叠成相同的二级结构，有 3 个茎环和 1 个单链区，此单链区类似于 U1 的 Sm 结合位点。因此 SL RNAs 存在和 snRNPs 相似的形式。可能作为 snRNP 中的一种。锥虫具有 U2、U4 和 U6 snRNAs，但没有 U1 和 U5 snRNA。U1snRNA 的缺乏可通过 SL RNA 的存在来解释，即 SL RNA 和 U1 snRNA 一样具有可以和内含子 5′端剪接位点识别和结合的功能。

SL RNA 的反式剪接反应可能表现了核内剪接装置的进化。SL RNA 提供了顺式剪接中的识别 5′剪接点的能力，可能依赖于 RNA 的特殊构象。它保留了剪接所必要的功能，在 mRNA 剪接中是由独立的 snRNAs 所提供的。无需像 U1 snRNA。SL RNA 这样的 RNA 就能执行剪接的功能表明 5′剪接位点的识别是直接依赖于 RNA。

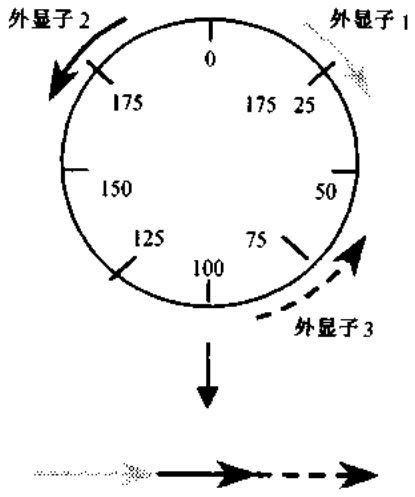


图 13－39　衣藻叶绿体的 *psa* 基因的反式剪接。（仿 B.Lewin: Genes Ⅵ, 1997, Fig. 30－23）

有的叶绿体基因也属于反式剪接的。图 13－39 表示衣藻叶绿体的 *psa*（photosynesizer Ⅰ）基因，它含有 3 个分散的外显子。外显子 1 离外显子 2 为 50kb，离外显子 3 为 90kb。很多别的基因位于这 3 个外显子之间。由于外显子 1 的转录方向和外显子 2 相反，即模板链不同，所以有时不能作为一个共同的转录本进行表达。其实，转录本是由单个外显子转录而成的。那么它们共处在一个 mRNA 上可能是通过 2 次反式剪接，连接而成。其他的反式剪接与此相同，mRNA 都是一个复合体。

第四节 植物病毒和类病毒的自我剪切

1986 年 Symosn 提出了**锤头二级结构模型**来解释某些类病毒，拟病毒和卫星 RNA 的自我剪切机制。

RNase P 的内切酶催化活性表现在催化外部的底物。RNA 具有内切酶活性的另一类情况是某些植物中小分子 RNA(约 350 个碱基)可以进行自我剪切反应。

这些小分子植物 RNAs 可分为两组：**类病毒**(viroids)和**拟病毒**(virusoids)，前者是感染性 RNA 分子，独自具有功能，外面不包被任何蛋白外壳。后者与此结构相同。但包被在植物病毒内。它和植物病毒的基因组一道被装配。拟病毒本身不能复制，必须依赖病毒的帮助，这种拟病毒有时称为**卫星 RNA**(satellite RNAs)。

类病毒和拟病毒两者复制都要经过滚环复制。RNA 链是被包被在病毒中，称为**正链**(plus strand)，在复制时产生的互补链称之为**负链**(minus strand)。现在发现了正链和负链的**多体**(multimers)。两种类型的单体可能是通过剪切滚环的尾巴而产生。环状的正链单体可能由一个线状单体末端相联而成(图 13－40)。

类病毒和拟病毒的正负链两者在体外都要经过自我剪切。此剪切反应是由二价阳

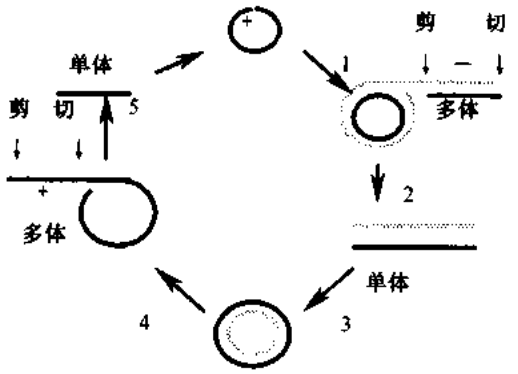


图 13－40 植物类病毒和拟病毒的复制与剪切。

离子启动的；它产生 5′－OH 和 2′－3′环磷酸二酯末端。这种 RNAs 有的在体外生理条件下剪切。而另一些都要先加热和冷却后再以同样处理才能剪切；这表明分离的 RNA 有一种不适当的构象，但经变性复性后能产生一种活性的构象。

大部分的类病毒和拟病毒的自我剪切取决于形成“**锤头**”(**hammerhead**)**二级结构**(图 13－41 上部)。这结构的序列对于剪切是充分条件。当周围序列缺失时，需要一个加热－冷却的周期，小分子 RNA 就能自发地自我剪切。此表明序列对活性的影响超过结构对活性的影响。

剪切的活性位点是仅有 58 nt 的序列。锤头结构含有 3 个茎环区，其位置和大小是固定的。另外还含有 13 nt 的保守序列，在此区域连结成结构的中心。这 13 个保守碱基和双链的茎形成了一个具有内在剪切能力的 RNA。

此结构的形成不需要通过内含子碱基配对。一个活性锤头结构也可以通过两条RNA 互补配对而构成。图 13－41 下部就是由 19 个碱基的 RNA 分子和另一个 24 个碱基的 RNA 分子杂交形成的锤头结构。这种**杂种 RNA 结构**和锤头结构很相似，就是少了Ⅰ环和Ⅲ环。当 19nt RNA 加到 24nt RNA 上时剪切就发生在锤头结构的适当位点上。

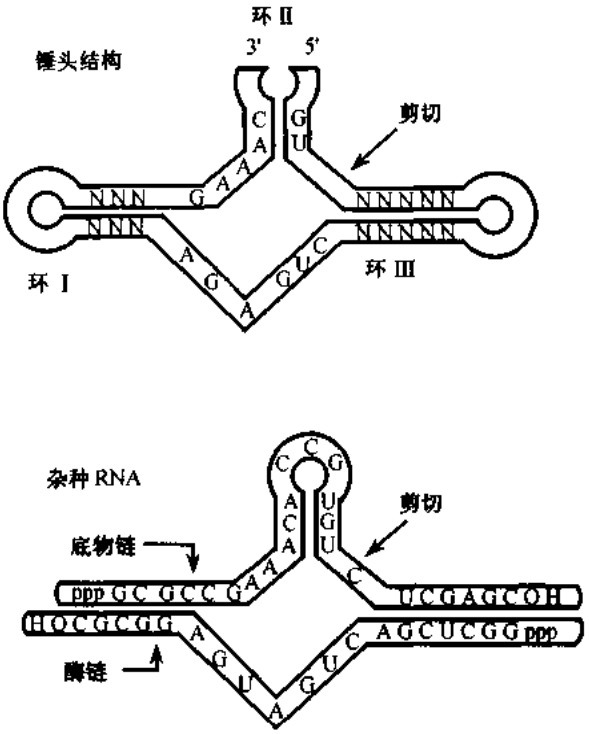


图 13－41　锤头结构和杂种 RNA 的自我剪切。(仿 B. Lewin: Genes V, 1994, Fig. 32－10)

我们将杂种 RNA 上部这条 24nt 的链看做是底物，下面 19nt 的链看成是酶。当把这 19nt RNA 和过量的 24nt 的 RNA 混合时，多拷贝的 24nt RNA 就会被剪切。这表明存在着一个 19nt～24nt RNA 配对，剪切和解离的循环(解离是指被切片段从 19nt RNA 中解离出来)。19nt RNA 再和新的 24nt RNA 配对。因此这个 19nt 的 RNA 是带有内切酶活性的核酶。此反应的参数和另一些 RNA 催化反应是相似的(表 13－3)。

表 13－3　各种 RNA 催化反应的参数

酶	底物	K_m	周转
19 b 类病毒	24bRNA	0.0006m mol/L	0.5/min
L－19IVS	CCCCC	0.04 m mol/L	1.7/min
RNase P RNA	Pre-tRNA	0.0005 m mol/L	1/min
RNase P 复合体	Pre-tRNA	0.0005 m mol/L	2/min
RNase T1	GpA	0.05 m mol/L	5700/min
β-半乳糖苷酶	乳糖	4.0 m mol/L	12500/min

引自 B. Lewin: Genes V, 1994, Table 32.2.

人们根据锤头结构的自体催化原理，设计了可进行异体催化的人工合成核酶。

Symons 等(1989 年)用 T7 RNA 聚合酶合成了紫花苜蓿短暂条纹病毒(VLTSV－A，一种拟病毒)RNA 正链序列中的一个 13nt 的 RNA(其作用相当于核酶)，和另一个 41nt 的 RNA(相当于底物)。这两个 RNA 在体外可以形成锤头结构。在镁离子存在下，前者可将后者剪切成 9b RNA(3′端为 2′－3′环化磷酸)和 32b RNA(5′端为羟基)，上述 13b 的 RNA 是人工合成最小的核酶。

Uhlenbeck 等(1987 年)合成了鳄型日斑类病毒(ASBV)的负链中的两个片段,在体外成功地进行了剪切反应。

Haseloff 和 Gerlach(1988)以氯霉素乙酰转移酶(CAT)mRNA 为对象,设计合成了锤头结构的核酶。在体外实验中,所有的锤头结构都可以切割 CAT 的转化产物。在此模式里,锤头结构所需的 13 个 nt 的保守碱基仅有两个来自底物,其余 11 个都来自人工合成的核酶,从而增加了核酶的应用价值。

近年来美国有几个实验室针对艾滋病毒的 *gag* 基因和 5'LTR(长末端重复序列)设计了四种核酶,每种核酶都能在体外生理温度下有效地切割靶 RNA,而其中一种核酶是在 Hela 细胞中表达出来的,因此他们认为核酶将有望应用于艾滋病的治疗。

Hampel 等 (1990) 将发夹型结构从分子内催化型转变成分子间催化型,并用突变和合成的方法推导出该结构中的必需序列和可变序列。他们还合成可切烟草环斑病毒卫星 RNA(STRSV),烟草斑纹病毒,CAT 和 HIV - 1(人类免疫缺病毒—即艾滋病毒)中 RNA 片段的核酶。在这种模式中,RNA 只要含有 GUC 即可作为底物,根据它周围的序列就可设计出相应的核酶。但剪切反应在底物识别信号(GUC)的 5'端,这和锤头结构不同。

第五节 编 辑

一、编辑的发现

编辑(editing)是指转录后的 RNA 在编码区发生碱基的加入,丢失或转换等现象。

1986.R.Benne 在研究锥虫线粒体 mRNA 转录加工时发现 mRNA 的多个编码位置上加入或丢失尿苷酸,1990 年在高等动物和病毒中也发现了编辑现象。

分子遗传学中一条基本原理就是 mRNA 的序列是由 DNA 中的编码区决定的。从 DNA 转录成 mRNA,从 mRNA 翻译成蛋白都体现了 1 对 1 的线性关系。断裂基因的发现和内含子的切除已使基因的表达增加了一个步骤:DNA 的编码序列(外显子)在 RNA 中必须重新连接。但这种加工仍然保持信息的传递。实际上 DNA 的编码序列依然保持其"纯洁",并未发生改变,也就是说虽然经过了加工,mRNA 所有的信息都来自 DNA。在哺乳动物和鸟类免疫球蛋白中可产生了新的编码序列,这是发生了一些特殊的行为,改变了 DNA 编码的信息,这就是免疫球蛋白的 DNA 体细胞重组。它的改变是发生在 DNA 水平的,而在 RNA 中并未产生外源的信息。RNA 的编辑就不同了,它是在 mRNA 的水平改变遗传信息的,而且往往会增加一些原来 DNA 模板中不曾编码的碱基。

RNA 的编辑因不同的原因会在不同的位置上发生。在哺乳动物细胞中,mRNA 中的个别碱基发生了置换,导致编码蛋白的氨基酸序列发生改变。在锥虫的线粒体中,几个基因的转录本发生了广泛的变化,碱基发生系列的增加和缺失。

图 13 - 42 表示载脂蛋白 - B(apolipoprotein-B, apo-B)基因,及其在哺乳动物肠和肝中的 mRNA。在各种组织的基因组中都鉴定出具有单个 *apo-B* 基因序列,在肝中翻译成 512kDa 的蛋白。而在肠中只翻译成 250kDa 的一个较短的蛋白。它仅有全长的一半(N 端)。经检测其 2153 位的密码子 CAA(Gln)中的"C"变成了"U",这样"CAA"也变成了终止密码子 UAA。

这一改变不是突变引起的。突变是指 DNA 上的碱基发生改变,而 *apo-B* 的 DNA 模

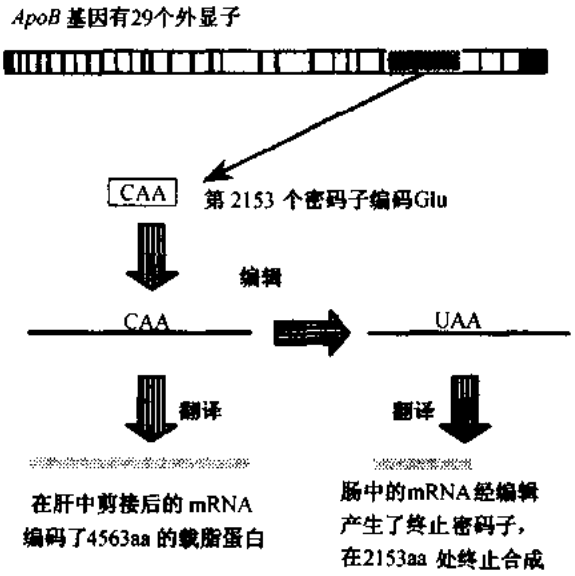


图 13－42　载脂蛋白的基因 *ApoB* 在肝和肠中的 DNA 序列是相同的，但在肠中经过编辑，引入了终止密码子，不能翻译成完整的载脂蛋白。（参考 B.Lewin: Genes Ⅵ, 1997, Fig. 31－14）

板并未发生改变；另外其发生的频率远高于突变。那么这种转换是怎样产生的呢？是转录本直接发生了变化。是否是由一种特殊的酶识别这个 *apo-B* 转录本，并将 C_{2153} 改变成 U 呢？这种酶可能以类似 tRNA 修饰酶的方式识别特殊的二级结构区域。经过建立了一种编辑的体外系统表明较小的序列（约 50 nt）在其编辑位点附近提供了一种指导序列。

这类编辑是很罕见的，但 *apo-B* 并不是惟一的例子。另一些例子是大鼠脑中的谷氨酸受体所提供的。编辑在单个位点上将原来编码 Gln 的密码子在 RNA 上变成 Arg 的密码子，而精氨酸在控制离子流向神经递质中起到重要的作用。因此编辑具有明显的生理功能。

在锥虫的线粒体的几个基因中发现了 RNA 序列有着戏剧性的改变。第一种情况是细胞色素氧化酶第Ⅱ亚基的蛋白涉及到 *cox Ⅱ* 基序列的－1 移框。图 13－43 表示

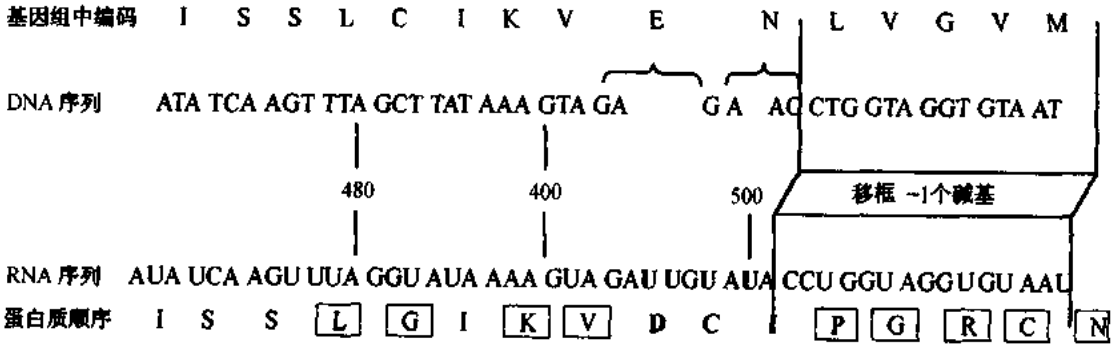


图 13－43　锥虫 *cox Ⅱ* 基因片断及其产物顺序的比较。核酸序列的数字是以起始密码子 AUG 的 A 开始编号。蛋白质以单个字母表示。方框表示与酵母、人类同源的氨基酸。黑体表示编辑位点。（参考 B.Lewin: Genes Ⅵ, 1997, Fig. 31－15）

DNA 的编码序列和所产生蛋白质的实际序列。这个基因有什么功能呢?

在 *cox* Ⅱ mRNA 中插入了 4 个核苷酸，插入的位点就在移框（frameshift）位点的边上。碱基的插入恢复了阅读框的特点，它插入了一个额外的氨基酸（以加入 4 个 nt 这个角度来说）并改变了其他位点的氨基酸。尚未发现别的基因有这种插入的情况。据推测额外的碱基是在转录后插入的。基因组的序列和 mRNA 序列出现差异的例子在 SV5 和麻疹副粘病毒（*measles paramyxoviruses*）的基因中也有所发现，在这些例子中涉及到在 mRNA 中加入 G 残基。

RNA 的编辑有时涉及很多碱基的加入和缺失。如在锥虫 *cox* Ⅲ 的 mRNA 上 158 位点插放了 394 核苷酸；在 9 个位点删去了 18 个尿苷酸，实际增加了 376 个核苷酸，使 *coxIII* 的长度增加了 55%。因此 *cox* Ⅲ 的基因比成熟的 mRNA 小了很多，故称其**隐匿基因**。

以上我们可以看出编辑具有着一定的生物学意义：①校正作用。因 U 的插入校正 *cox* Ⅱ 的移码，使其蛋白质的序列和其他生物的相似。②调控翻译。通过编辑可以构建或去除起始密码子和终止密码子。例如 *apo - B* 基因在肠中表达时通过编辑，引入 UAA，使合成到一半就终止了。③扩充遗传信息。锥虫 *cox* Ⅲ mARNA 编辑后增长了 55% 等。

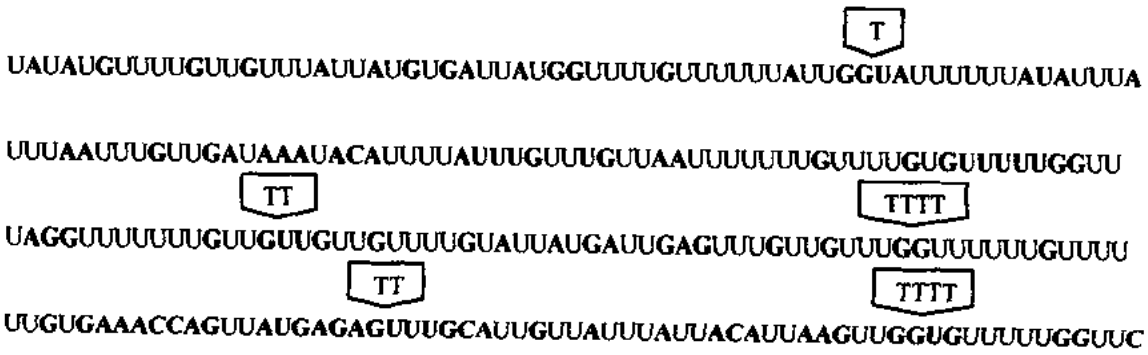


图 13 - 44 锥虫（*T. brucei*）*cox* Ⅲ 基因的部分 RNA 顺序。很多 U（非黑体）在 DNA 中未编码，而另一些在 DNA 中编码的 T（用灰色表示）在 mRNA 中被删除了。（参考 B. Lewin: Genes Ⅵ, 1997, Fig31 - 16）

二、编辑的机制

是谁为尿苷的插入以及碱基的缺失提供信息呢？1990 年 L. Simpsom 等在研究锥虫线粒体 mRNA 时发现了一类新的小分子 RNA，这种 RNA 可以和 mRNA 分子被编辑的部分发生非常规的 G - U 配对互补，对 mRNA 前体分子的编辑起了指导作用，故称其为**指导 RNA**（guide gRNA）。gRNA 含有一段序列可和被编辑的 mRNA 互补。图 13 - 45 表明利什曼原虫（*Leishmania*）的细胞色素 B 基因（*CoB*）中 gRNA 的作用。上面的是原来的转录本，或者被编辑的前体 RNA，空缺的位置是将要通过编辑指放碱基的位置。gRNA 可和被编辑的前体 RNA 分子的编辑区及相邻的序列互补结合。3′端的互补区要长于 5′端。在互补区中 gRNA 的 A 的相应位置留下了空缺。gRNA 提供了 U 插入位置的模板。当反应完成时 gRNA 和 mRNA 彼此分离，mRNA 已成熟，可以进一步进行翻译。

gRNAs 是作为一种独立的转录单位被编码的。图 13 - 46 表明利什曼原虫线粒体

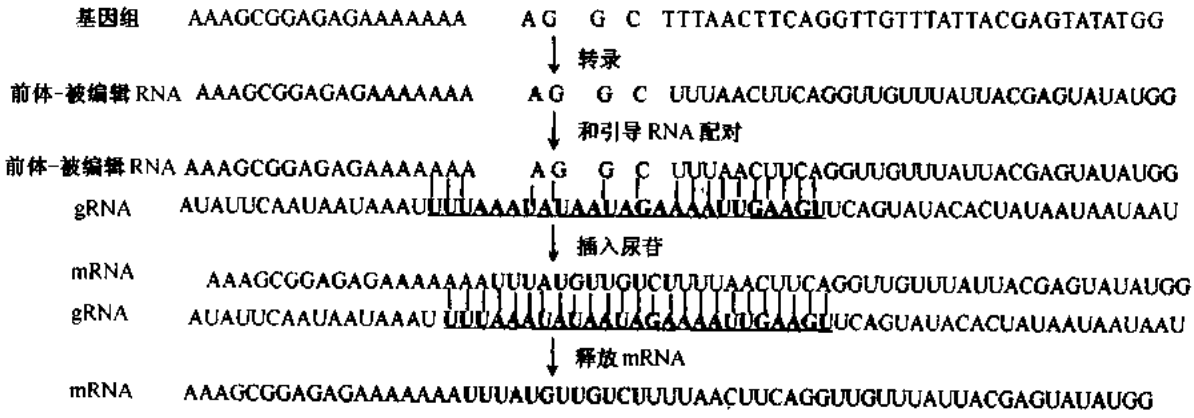


图 13－45　被编辑的前体 RNA 和引导 RNA(gRNA)配对，以 gRNA 为模板插入尿苷，完成编辑。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 31－16）

DNA 相关区域的图。此区含细胞色素（cyB）基因，它编码了被编辑的前体 RNA 序列和 2 个特殊的 gRNA。主要编码区的一些基因和它们的 gRNA 是分开的。原则上无论是基因突变还是 gRNA 发生突变都可能导致 mRNA 和蛋白质一级序列的改变。根据遗传学标准，每一个单位可能被认为由基因的一部分组成。由于这些单位可以独立地

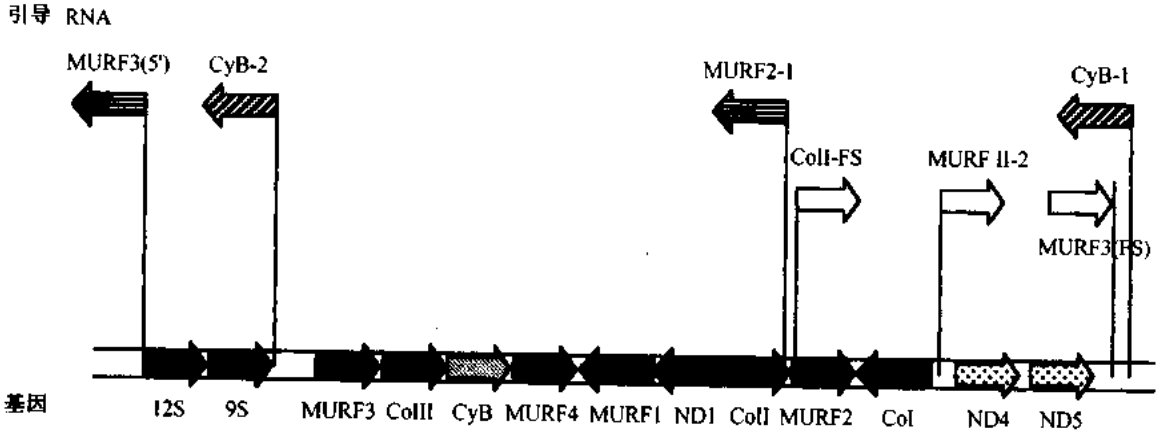


图 13－46　利什曼(*Leishmania*)原虫编码被编辑的基因和它们的 gRNA 基因是分散排列的，有的基因有多个 gRNAs。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 31－18）

表达，应属于反式互补。通过突变我们可以发现 3 个互补群对于编码单个蛋白的一级序列是必要的。

转酯反应机制也是一个十分引人注目的问题。1990 年 Blum, B 等首先提出了**编辑体模型**；1991 年 Cech 等又提出了以 gRNA 和被编辑前体 RNA 配对，由寡聚 U 进行转酯反应的模型。同年 Blum 等又提出了 RNA 编辑的**转酯模型**(图 13－47)。

通过部分编辑的研究表明加工的方向是延着被编辑的前体 RNA(pre-edited RNA)的 3′－5′进行。引导 RNA 不仅指导尿苷通过和被编辑的前体 RNA 的配对插入到特异位点。同时也提供了插入尿苷的来源。

引导 RNAs 末端延伸着 3～5 个 U 残基。U 苷在事前被加到 gRNA 3′端寡聚 U 的尾

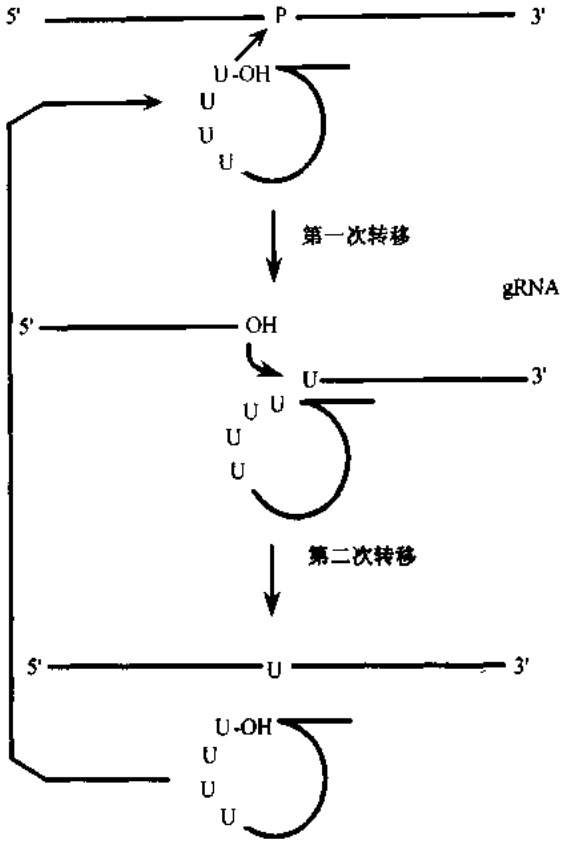


图 13-47 RNA 编辑的转酯反应模型。(仿 B.Lewin: Genes Ⅵ,1997,Fig.31-19)

巴上。这个尾巴的长度是异质性的,从 5 个到 24 个 U 不等。这些 U 通过类似于 1 类内含子的剪接反应插入到被编辑的 RNA 中。gRNA 和被编辑的前体 RNA 的配对时也类似于内含子内部引导序列(IGS)和 5′外显子的配对。图 13-45 表示编辑过程中的二步转酯反应。

第一次转酯是编辑区与 gRNA 配对,gRNA 聚寡 U 尾巴的 3′-OH 向编辑区作亲核进攻,使被编辑的前体 RNA 切开为二部分,一部是 5′区域,其末端带 3′-OH,另有一部分是 3′端部分,它和 gRNA 的 5′端配对,其 5′端和 gRNA 的 3′端形成了新的磷酸二酯键,也像是个套索。但并不存在分枝结构。实际上中间体是线形的,仅 5′端和被编辑的前体 RNA 编辑区邻近区配对形成了环。

第二次转酯反应是第一次转酯反应产生的 3′-OH 对 gRNA 尾部区域的 2 个 U 残基之间进行亲核进攻,这样被编辑的前体 RNA 的 5′区和 3′区又重新连接起来,恢复原状,但在编辑位点已插入了一个 U 残基。这样再照此往返重复,加入更多的 U。

被编辑分子的结构表明,每次反应只加入 1 个 U,而不是加入一组 U。很可能编辑的加工是通过连续的循环加上 U 残基。现在尚不知道 U 的缺失和 C 的加入等是否也是通过与上述相似的反应。

第十四章

遗传密码和遗传信息的翻译系统

基本概念

遗传密码是三联密码，在一个 mRNA 中 3 个核苷酸形成一个密码子，编码一个氨基酸。有的氨基酸是由一个以上的密码子编码的。遗传密码是通用的，阅读时没有间隔，且是连续的，不重叠的。

mRNA 在核糖体上翻译成肽链。氨基酸由 tRNA 分子运送到核糖体上。正确的氨基酸序列是通过 mRNA 上的密码和 tRNA 上的反密码子互补结合来完成的，而每个特异的氨基酸又是特异地结合到 tRNA 上。

在原核和真核生物中，AUG(甲硫氨酸)是翻译开始的起始密码子。肽链的延伸涉及到结合在核糖体的 A 位和 P 位上的 tRNA 所携带的氨基酸之间形成肽键。一旦肽键形成，核糖体沿着 mRNA 移动一个密码子，为下一个带着氨基酸的 tRNA 的结合做好准备。

翻译是持续的，直到 mRNA 上终止密码子(UAG，UAA 或 UGA)子的出现方完成。这些终止密码子是由一个或多个释放因子蛋白来阅读，然后将多肽从核糖体上释放出来，蛋白质合成装置的其他成份也将解离。

在真核生物中，蛋白质有些游离在细胞质中的，有些存在于各种细胞器中，如核、线粒体、叶绿体和分泌泡中。预定要进入细胞器的蛋白质具有特殊的氨基酸序列，这些序列指导蛋白质越膜，到达不同的位置。如分泌蛋白具有 N 端的信号序列，使它们能进入粗面内质网，然后再进入高尔基体，最终分泌到体外。预定要到核、线粒体或叶绿体的一些蛋白，也都有特异序列，使它们最后定位于这些细胞器中。

我们前面讨论过作为遗传物质必须要具备三个条件(见第 9 章)，其中之一就是必须含有一个有机体的结构，功能，发育和繁殖的全部稳定的信息。生物的所有这些特点都要涉及到特异蛋白质的作用，有关蛋白质的信息是由结构基因编码的。编码蛋白的基因的表达主要发生在两个阶段：转录和翻译。**转录**是将 DNA 上的遗传信息转变成单链 mRNA 分子的碱基序列。这个碱基序列携带着多肽的氨基酸序列的特殊信息。在细胞中 mRNA 的碱基序列信息转变成多肽的氨基酸序列称为**翻译**(translation)。决定多肽特异氨基酸序列的 DNA 碱基的信息就称为**遗传密码**。

首先我们来介绍一下遗传密码的破释，这也是遗传学发展中的重大事件之一。遗传密码的破译，测序方法的建立以及体外重组的实现是基因工程的三大基石，因此在遗传密码破译后仅 2～3 年的时间这项工作就获得了诺贝尔奖，由此可见意义之深远，功绩之巨大。

第一节　遗传密码的破译

自从“一基因一酶”学说建立(1941 年)以后，人们逐步地认识到基因和蛋白的关系。“中心法则”提出后更为明确地指出了遗传信息传递的方向，总体上来说是从 DNA→RNA→蛋白质。那 DNA 和蛋白质之间究竟是什么关系？或者说 DNA 是如何决定蛋白质？这个有趣而深奥的问题在五十年代末就开始引起了一批研究者的极大兴趣。早在 1944 年理论物理学家 Erwin Schrödinger 发表的《什么是生命》一书中就大胆地预

言，染色体是由一些同分异构的单体分子连续组成。这种连续体的精确性组成了遗传密码。他认为同分异构单体可能作为一般民用的莫尔斯电码的两个符号：“·”、“—”，通过排列组合来储存遗传信息。此时遗传物质的化学本质尚未确定，同年 Avery 虽成功地完成了体外转化实验，但尚未改变人们认为蛋白质是遗传物质的传统观念。十年后 DNA 双螺旋模型才得以建立，在这样的背景下能将遗传信息设想成一种电码式的遗传密码形式，实在是一种超越时代的远见卓识。到五十年中期双螺旋模型的建立，给予科学家们以很大的激励。破译遗传密码也就成了势在必行的工作。

要破译一个未知的密码，一般的思路就是比较编码的信息，即密码和相应的译文。对于遗传密码来说最简单的破译方法应是将 DNA 序列或 mRNA 序列和多肽相比较。但和一般破译密码不同的是，遗传信息的译文——蛋白的序列是已知的，未知的却是密码。1954 年 Sanger 用纸层析分析了胰岛素的结构后，对蛋白质的氨基酸序列了解得越来越多。但是直到 1965 年前后经历了十年时间，多位科学家的执着研究才破译了密码，其中最为重要的几项工作其思路之新颖方法之精巧都闪烁着科学的智慧之光。

一、遗传密码的试拼

1954 年科普作家 G. Gamov 对破译密码首先提出了挑战。他以著有《奇异王国的汤姆金斯》等优秀的科学幻想作品而著称，具有丰富的想象力，但他不是一位实验科学家，所以只能从理论上来尝试密码的解读。是年，他在 *Nature* 杂志首次发表了遗传密码的理论研究的文章，指出“氨基酸正好按 DNA 的螺旋结构进入各自的洞穴”。他设想若一种碱基与一种氨基酸对应的话，那么只可能产生 4 种氨基酸，而已知天然的氨基酸约有 20 种，因此不可由一个碱基编码一种氨基酸。若 2 个碱基编码一种氨基酸的话，4 种碱基共有 $4^2 = 16$ 种不同的排列组合，也不足以编码 20 种氨基酸．因此他认为 3 个碱基编码一种氨基酸的就可以解决问题．虽然 4 个碱基组成三联密码，经排列组合可产生 $4^3 = 64$ 种不同形式，要比 20 种氨基酸大两倍多，但若是四联密码，就会产生 $4^4 = 256$ 种排列组合。相比之下只有三联体(triplet)较为符合 20 种氨基酸。后来的实验证实这一推测是完全正确的。但人们不禁要问在三联体中的每个碱基作为信息只读一次还是重复阅读呢？Gamor 也许是考虑到效率的问题，认为一个碱基可能被重复读多次，也就是说遗传密码的阅读是完全重叠的，因此氨基酸数目和核苷酸数目存在着一对一的关系。这一假定非常简洁地解释了核苷酸间距和多肽链上邻接氨基酸的间距(0.36nm)之间显示了明显的相关性。若真如此，重叠密码对多肽链上氨基酸的序列就形成了一种限制。例如，具有完全重叠密码的密码子 ATC，后面接着的密码子一定是 TC 开头，那么相应的氨基酸的序列也会受到限制。再者若是重叠密码，那么任何一个碱基的突变都会影响到相连的 3 个重叠密码子，即三个氨基酸都会发生改变，但事实并非如此。1957 年 S. Brenner 发表了一篇令人兴奋的理论文章，他通过蛋白质的氨基酸序列分析，发现不存在氨基酸的邻位限制作用，从而否定了遗传密码重叠阅读的可能性。同时人们也发现在镰刀形细胞贫血的例子中，血红蛋白中仅有一个氨基酸发生改变。说明 Gamov 的后一推论是错误的。这就是智者千虑，必有一失。很多著名的科学家也有过类似的失误。在资料较少的情况下，对未知的真理作出推断，难免会发生偏差，但瑕不掩瑜，人们对他们的那种敏锐、大胆、睿智和创新的精神，巧妙的构思仍敬佩不已。

二、三联密码的验证

三联密码是否真实存在？若真实存在那么共有 64 个密码子，而氨基酸只有 20 种，那么多余的 44 种有何用处。1957 年 Crick、Griffith 和 Oregel 等为了解释这个问题提出了一个设想。首先认为象 AAA，GGG，CCC，TTT 这四个三联体，分别由相同的碱基构成，解读的起始位置有可能发生差错，因此可能是"无义"密码子。这样余下的只有 60 个密码子。接着他们又设想，例如 ATT 和 GCA 若分别编码氨基酸 a 和 b，若这两个密码子连续排列成 ATTGCAATT……，在起读时若发生错位就会产生 TTG，TGC，CAA 和 AAT 等序列就是错读，这些错读的重迭密码也是无意义的，也就是说一个序列有 3 种读法，其中只有一种是有意义的，而其余的两种都是无义密码，这样（$60 \times 1/3 = 20$）有义密码子只有 20 个，似乎是很圆满地解释了氨基酸数目和密码子总数之间的矛盾，但后来的实验证明，此设想也是重蹈 Gemov 的覆辙。

直到 1961 年 Crick 和 Brenner.S 等设计的实验有力地证实了三联密码的真实性。他们用 T4 染色体上的一个基因（*r* Ⅱ 位点）通过用原黄素（proflavin）处理，可以使 DNA 脱落或插入单个碱基，插入叫**"加字"突变**，脱落叫**"减字"突变**（图 14－1），无论加字和减字都可以引起移码突变。Crick 小组用这种方法获得一系列的 T4"加字"和"减字"突变，再进行杂交来获得加入或减少一个、二个、三个的不同碱基数的系列突变。

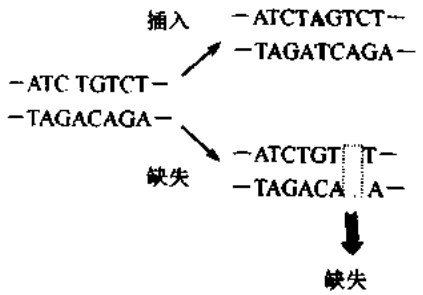


图 14－1　"加字"和"减字"突变。

开始用原黄素诱导的突变称 FCO，它们只能在 B 菌株上生长形成噬菌斑，而不能在 K 菌株上生长，然后他们再用原黄素诱导产生回复突变，在 *E.coli K*(λ)菌株中出现了噬菌斑（见第 6 章顺反子）。用遗传学的方法和野生型杂交（图 14－2）分析这些"回复突变体"，发现它们并不是真正的野生型。表明此回复突变并非是突变位点又精确地回复到原来的状况，实际上这种所谓回复突变是在不同位点发生第二次突变而引起的。这种第二次突变"抑制"了原来 FCO 的表达。**校正突变**（A suppressor mutation）抵消或抑制了前一次突变的效应，校正突变的特点如下：

（1）校正突变是在第一次突变不同位点将它抵消的。因此原来的突变可以通过野生型和回复突变型之间的杂交又恢复为突变型。

（2）校正突变可能发生相同的基因中，抑制原来的突变（如刚举的例子），称基因内抑制，或发生在不同的基因中称**基因外抑制**。

（3）不同的抑制可能作用的方式不同。如有的抑制是在转录和翻译水平，有的可能是通过细胞的生理功能来实现。

图 14-2 表示用原黄素诱导的突变型 FCO 经再次诱导在同一顺反子内的别的位点上产生新的突变，此突变抑制了原来的突变，使 FCO 得到恢复。若将这种恢复突变型和野生型进行杂交，由于染色体交换的结果又产生了 FCO 和另一种突变型。

怎样来解释这些结果呢？我们可以推测阅读是有极性的，也就是从一端开始阅读。第一次由原黄素诱导的插入或缺失可能会干扰正常的阅读机制，就是一组碱基读成一个字，例如 mRNA 上的 3 个碱基作为一个**阅读框**(reading frame)，头 3 个就读成第一个字。当原黄素诱导使 DNA 上的增加或减少了单个碱基将会使阅读框移动，导致"错误"，这样的移框突变可能导致遗传信息变成"无义"的。若第二次诱变相应地插入或缺失一个碱基可以恢复正确的阅读框。

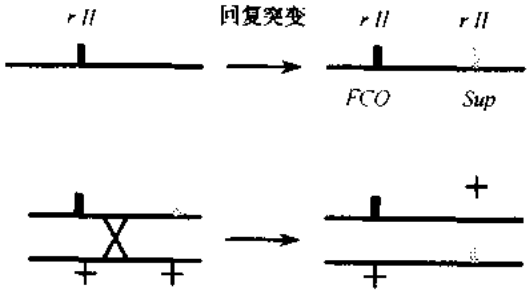


图 14-2 抑制基因所产生的回复突变型和野生型杂交后，经交换可产生原有的突变型。
(仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 13-16)

通过这样的方法他们发现加入或减少一个和二个碱基都会引起移码突变，而加入或减少 3 个碱基时反而可以恢复正确的读框，表明每个密码的确是由 3 个碱基组成的。

Crick 对遗传密码提出了 4 个特点：①3 个碱基一组，编码一个氨基酸(是通过本实验所证实的)；②密码是不重叠的（实验证据来自 Wittmann 等的实验及 A.Tsugita, H.Fraenkel-Conrat 用亚硝酸诱发烟草花叶病毒 TMV 产生的突变体的研究中得到的)；③碱基的序列是从固定起点解读的，即 mRNA 具有固定的阅读框；④密码是简并的，即某个特定的氨基酸可以由几个碱基三联体来编码。否定了他们以前的解释，即 64 种密码子中只有 20 种编码，其余的 44 种都有是无意的这一推测。从他们的实验结果来看，如果以前的解释是正确的话，那么任何移码突变都将是无意突变，那么 FCO 系统突变体的那个区域(约为顺反子 B 的 1/4)应当很小，但其实不然，发生移码仍可翻译，只不过肽链的序列发生很大的改变，而不是产生很短的肽链。

三、利用突变来解读密码

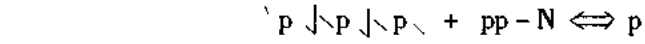
A. Tsugita, H. Fraenkel-Conrat 小组和 H.G. Wittmann 小组试图通过用亚硝酸来对 TMV 进行诱变。当时（1960 年）已搞清了 TMV 肽链的一级结构由 158 个氨基酸组成，将突变型和野生型进行比较就能确定肽链上氨基酸取代的位点和类型。当时根据亚硝酸诱变的原理，推测诱变的产生必定是由于 mRNA 中的 A→G 或 C→U 的缘故。那么取代与被取代的氨基酸之间，一定有两个碱基是相同的，而第三个碱基原来的氨基酸应是 A 或 C，也有可能 A，C 都有，而取代的氨基酸的第三个碱基应为 G 或 U。他们一共获得了约 200 个突变株，经反复比较分析结果破释了少数几个密码子，离全部破译

尚有很大的距离。但他们直接地证实了密码子是不重叠的。若当时可以测序的话，用这一方法可以破译所有的密码子。

四、无细胞系统的建立

1961 年 Nirenberg 建立了无细胞系统，为遗传密码的破译和以后的体外表达奠定了重要的基础，这一新技术又是在多核苷酸磷酸化酶发现的基础上建立起来的。

1955 年也就是在 RNA 聚合酶发现的前 5 年，Marianne Grunberg - Manago 和 S.Ocha 在细菌中分离了在体外合成多核苷酸中起作用的另一个酶，定名为**多核苷酸磷酸化酶**（polynucleotide phosphorylase），它催化核糖核苷二磷酸的聚合，反应如下：



多核苷酸磷酸化酶不需要任何 DNA 模板就可合成，这是和 RNA 聚合酶及 DNA 聚合酶有所不同。

在美国国立卫生研究院（NIH）从事研究工作的青年科学家 Nirenberg.M.W 在读到第一篇发现 mRNA 的报导之后，就决定计划建立一种无细胞反应系统，来揭开遗传密码之谜。

他们的方法是：

（1）去模板：用 DNAase 处理 *E.coli* 抽提物，使 DNA 降解，除去原有的细菌模板。在抽提物中含有核糖体，ATP 及各种氨基酸，除 mRNA 以外，是一个完整的翻译系统。由于 DNA 被降解，所以不再转录新的 mRNA，即使原来残留的 mRNA 因其半衰期很短，也很快会降解掉。

（2）加入 pol U：Nirenberg 成功地破坏了翻译系统中的内源 mRNA，这样从理论上来说若加入任何外源 mRNA 就可以按新的信息合成蛋白。他们采用了多核苷酸磷酸化酶，仅以尿苷二磷酸为底物，人工合成 poly U。当他们把人工合成的 polyU 加入这种无细胞系统中代替天然的 mRNA 时，惊喜地发现果真合成了单一的多肽，即多聚苯丙氨酸，它的氨基酸残基全是苯丙氨酸，这一结果不仅证实了无细胞系统的成功，同时还表明 UUU 是苯丙氨酸的密码子。他们用同样的方法分别加入 polyA，polyA 和 polyG 结果相应地获得了多聚赖氨酸，多聚脯氨酸和多聚甘氨酸。Nirenberg 利用无细胞系统不仅顺利地破译了 4 个密码子，同时又证实了 Crick 等原先认为 AAA，UUU，GGG，CCC 是无义密码子的推测是错误的。

（3）按比例加入 2 种核苷混合的多聚物。

由于当时还未分离 RNA pol 酶，无法按设计的模板来合成 RNA，但除了 UUU，CCC，AAA，GGG 以外，还必须破译其他的密码，Nirenberg 又想出了一种新的方法，就是按一定的碱基比例来合成 RNA。比如在底物中加 5 份的 UDP 和 1 份的 GDP，碱基比为 U：G = 5：1，它们能组成的三联体不外乎 8 种：UUU，UUG，UGU，GUU，GGG，GGU，GUG，UGG。U 和 G 将随机地加入到三联体中，这样按比例各个位于上进入 U 和 G 的概率不同，如 UUU：UGG = (5 × 5 × 5)：(5 × 1 × 1) = 25：1；同理 UUU：UUG = 5：1，根据这样的推测，在无细胞系统中以这种比例合成的 mRNA 产生的氨基酸的比例也应是相应的，这样可以推测出密码子的组成。如氨基酸测定结果：

苯丙氨酸(UUU)：半胱氨酸(UGU) = 5：1

苯丙氨酸(UUU)：缬氨酸(GUU) = 5：5

苯丙氨酸(UUU)：甘氨酸(GUU) = 24：1

苯丙氨酸的密码子是已知的，由 3 个 U 组成，那么半胱氨酸一定是由 2 个 U，1 个 G 组成；缬氨酸同样如此；甘氨酸应是由一个 U 两个 G 组成。S. Ochoa 及其合作者获悉 Nirenberg 用 polyU 大获成功之后，利用身边保存着多种多聚核苷酸也开展了破译密码的研究，采用的方法也是加入不同比例的混合多核苷酸。两组展开了激烈的竞争，经过两个组一年多的努力，结果搞清了各种氨基酸的碱基组成，但是并不知其序列。Nirenberg 于 1964 年又采用三联体结合实验，一举破译了所有密码，取得了重大的突破。

五、三联体结合实验

从上面的实验结果不难看出，按比例合成 RNA 的方法不能解决最关键的序列问题，此时擅长 RNA 合成的 G. Khorana 就担负起直接合成有序多核苷酸的难题，1964 年正当 Khorara 刚刚奋力完成了第一批排列的核糖多核苷酸时，Nirenberg 又有新的突破，使破译密码的艰难工作迅速达到了光辉的顶点，这种新的突破就是建立了三联体结合

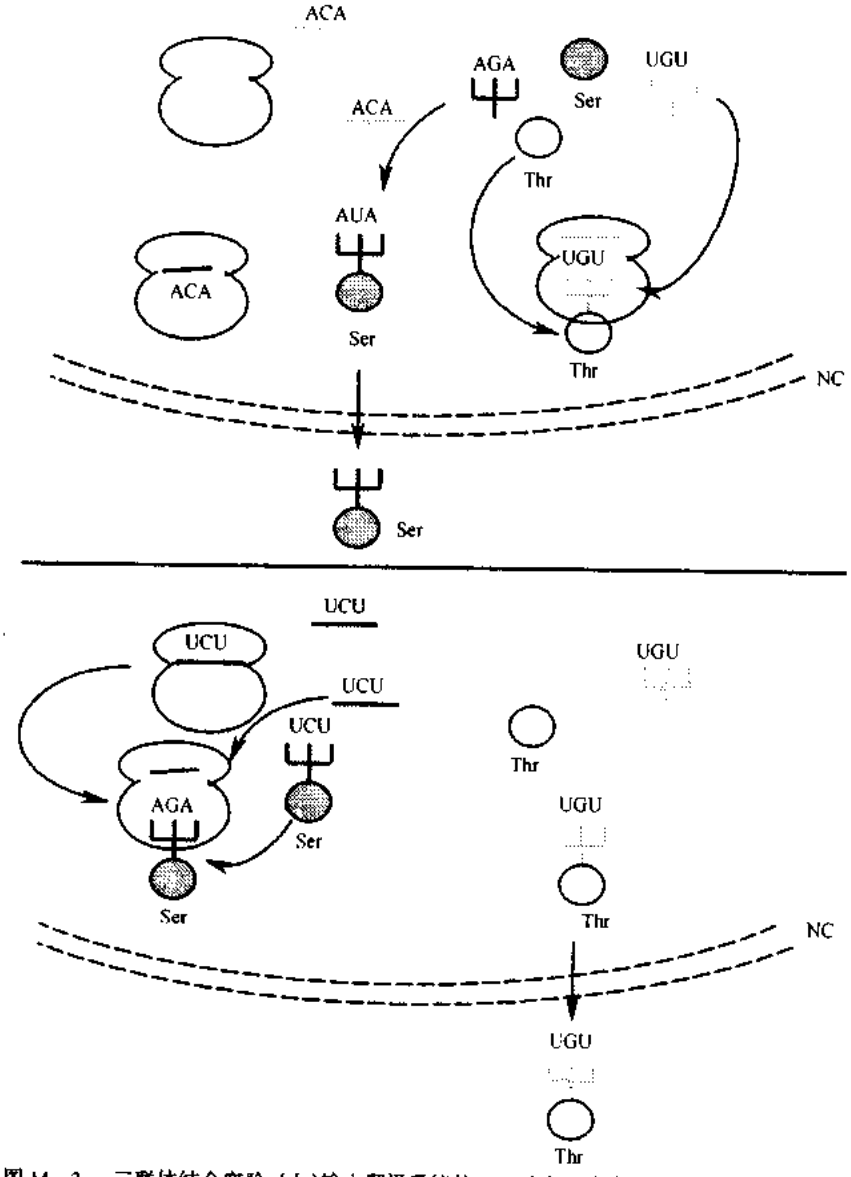


图 14-3　三联体结合实验。(上)输入翻译系统的 RNA 和标记的氨基酸(以灰色表示)不符。(下)输入翻译系统的 RNA 和标记的氨基酸(以灰色表示)相符。

的新方法。这个方法的思路是建立在两项基础上的：① tRNA 和氨基酸及三联体的结合是特异的；②上述结合的复合体大分子是不能通过硝酸纤维滤膜的微孔，而 tRNA－氨基酸的复合体是可以通过的。

Nirenberg 采用了一把钥匙开一把锁的思路，进行破译密码。他们首先发现当简单的特定的核苷酸加入到 *E. coli* 的核糖体上时，它们并不促使蛋白质的合成，而引起了特定的 tRNA 及其携带的氨基酸结合到核糖体上，形成大的复合体。因此他们每次在无细胞系统中仅加一种已知序列的三联体 RNA(如 ACA)，同时在氨基酸中只用 ^{14}C 标记一种氨基酸(如 Ser)，若 ACA 进入核糖体后，与其结合 tRNA 上携带的不是所标记的 Ser，那么 $tRNA^{Ser}$ 和其携带的 Ser 就会从 NC 上透过，所以通过测定透过 NC 的 tRNA－aa 复合体是否带有标记，如带有标记就可以确定输入的三联体 ACA 不是 Ser 的密码子；那么就可重新输入另外的三联体 RNA，一直到 tRNA 所带有的标记的氨基酸不透过 NC，说明此三联体 RNA 正好是标记氨基酸的密码子(图 14－3)，就这样 Nirenberg 小组一举破译了全部的密码。

六、利用重复共聚物破译密码

1965 年 Khorana 以不同的思路和方法也巧妙地破译了全部的密码，他发挥了自己合成 RNA 的特长，用已知碱基组成二个、三个或四个碱基合成重复序列的 mRNA，在体外翻译系统中加入同位素标记的氨基酸，然后分析所合成多肽的氨基酸序列，再进行比较分析。

Khorana 采用了有机合成一条短的单链 DNA 重复序列，然后用 DNA pol Ⅰ合成其互补链，再用 RNA pol 及不同的底物合成两条重复的 RNA 共聚物(图 14－4)，作为被翻译的 mRNA ，加入到体外表达系统中，根据合成的肽链(以同位素标记)的相应序列来推测各氨基酸的密码子。如表 14－1 所示，当重复序列为(UC)n 时，组成的重复 RNA 无论怎么阅读，只可能是 UCU－CUC，翻译的多肽也是由丝氨酸和亮氨酸之间排列的序列，但尚不能确定这两种氨基酸的相应密码子。当重复序列为 $(UUC)_n$ 时，无论怎么阅读，都只产生三种多聚氨基酸，即 poly Ser，poly Leu 和 poly Phe，和第一次比较，只有一个密码子 UCU 相同，但同样都有 Ser 和 Leu，所以仍不能确定。再看第三行重复序列(UUAC)n，无论怎么读法，只会是四个密码子的循环；UUA-CUU-ACU-UAC，但合成的肽链中氨基酸三种，-Leu-Leu-Thr-Tyr 。将密码子和氨基酸与第二次作对照，彼此共有密码子 CUU 和 Leu，所以可以确定 CUU 是 Leu 的密码子。那么第二栏中既然 CUU 已知是亮

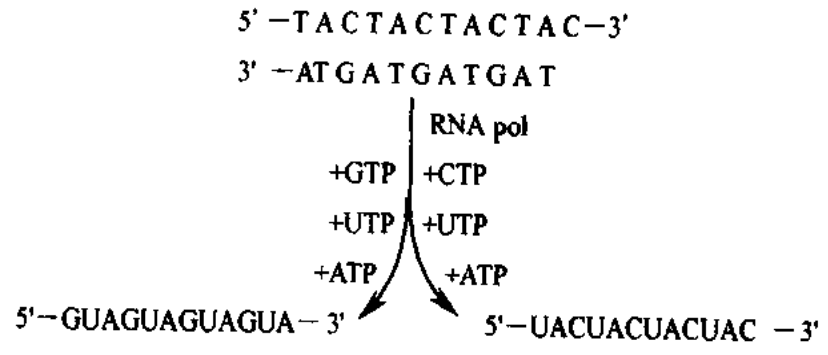


图 14－4　重复共聚体 RNA 的合成途径。(仿 Watson, J.D. *et al.*: Molecular Biology of the Gene, 4th ed., 1987, Fig. 15－4)

氨酸，毫无疑问 UCU 是丝氨酸。第一栏中原来 UCU－CUC 难以确定那一个是 Ser，那一个是 Leu，现已确定 UCU 是 Ser，那么余下的 CUC 定是亮氨酸了。Khorana 就用这种方法将所有的遗传密码都破译了。这项实验还同时证实了三联密码的正确性，以及简并的存在。由于 Nirenberg 和 Khorana 二人在破译遗传密码研究中的卓越贡献，他们二人共同获得了 1968 年的诺贝尔化学奖。

表 14－1　用二个或三个、四个核苷酸构造重复共聚体来确定密码子

重复顺序	可组成的三联密码	多肽的氨基酸组成
(UC)n	UCU-CUC	Ser-Leu
(UUC)n	(UUC); (UCU); (CUU)	poly Phe, poly Ser, poly Leu
(UUAC)n	(UUA-CUU-ACU-UAC)	Leu-Leu-Thr-Tye

表 14－2　遗传密码

第一位	第二位 U		第二位 C		第二位 A		第二位 G		第三位
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC		UCC		UAC		UGC		C
	UUA	Leu	UCA		UAA	终止	UGA	终止	A
	UUG		UCG		UAG	终止	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	UGU	Arg	U
	CUC		CCC		CAC		GUC		C
	CUA		CCA		CAA	Gln	UGA		A
	CUG		CCG		CAG		UGG		G
A	AUU	Ile	ACU	Thr	AAU	Asn	UGU	Ser	U
	AUC		ACC		AAC		UGC		C
	AUA		ACA		AAA	Lys	UGA	Arg	A
	AUG	Met	ACG		AAG		UGG		G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

七、终止密码子的确定

1964 年 Yanofsky 在研究 *E.coli* 色氨酸合成酶 A 蛋白时预言无义密码子的存在。他的推测是从两个不同的角度出发的：一是为 trp A 编码的 mRNA 还编码了 trpB，trpC，trpD 和 trpE。即一个 mRNA 分子中可以作为不同多肽的模板，那么有可能在翻译时中途在某个位点（两个肽的连接处）停止，然后再从下一个新的起点翻译，这样使各个肽可以分开，而不至于产生一条很长的肽链。这就意味着**终止密码子**的存在。另外他发现 *E.coli* Trp^- 的突变株是不能合成完整的色氨酸合成酶蛋白，但继续对它进行诱变可以得到回复突变。回复突变有两种，一种是个别氨基酸发生了变化，而另一种是完全回复，没有任何氨基酸组成的变化，这表明，*E.coli* Trp^- 不可能是任何移码突变的结果，那么这类的突变很可能携带有阻止合成的无义密码子。

1962 年 Benzer 和他的学生 S. Champe 对 T4 *r* Ⅱ突变的研究中发现野生型的 T4 *r* Ⅱ有两个顺反子 *r* Ⅱ *A* 和 *r* Ⅱ *B*，共同转录一个多顺反子 mRNA，但翻译成两个分开的蛋白 A 和 B。当发生缺失突变时，其中有一个突变型为 *rl* 589，证明是缺失所造成，缺失的区域含 *r* Ⅱ *A* 基因右边的大部分，和 *r* Ⅱ *B* 左边的小部分。互补实验表明 *rl* 589 的产物是一条多肽，但无 A 蛋白的活性，但有 B 蛋白的活性。Benzer 认为，这种缺失可能使 mRNA 失去了 A 蛋白合成"终止"和"B"蛋白合成"起始"的密码子，因此翻译时沿着一条 mRNA 阅读下去，产生了一条长的肽链(图 14－5)。

1964 年 Brenner 及其同事获得了 T4 噬菌体编码头部蛋白基因的琥珀突变(amber)，进行了精细作图(图 14－6)，并分离研究了各种突变型的多肽。突变型的肽链比野生型的要短，因此可以推测琥珀突变可能产生终止密码子，使肽的合成在中途停止下来；由于突变位点越靠近基因的左端，所产生的肽链越短，越靠近右端越接

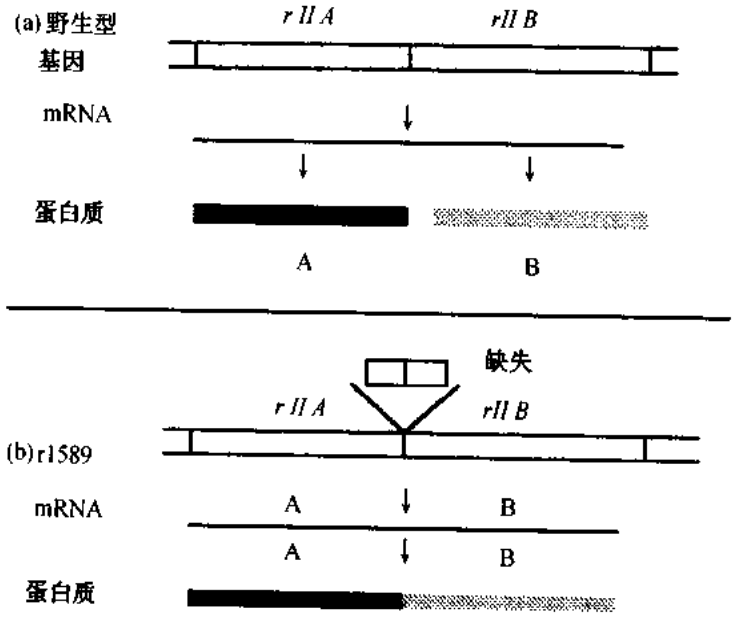


图 14－5　噬菌体 T4 的 *r*1589 缺失突变型。(a)野生型 *r* Ⅱ *A*，*r* Ⅱ *B* 产生多顺反子 mRNA，但翻译成两个分开的蛋白 A 和 B。(b)可能丢失了 A 基因末端的终止密码子和部分编码区，结果产生了没有 A 活性的单个融合蛋白分子。

近野生型，据此可以推测翻译的过程是从 mRNA 的 5′端向 3′阅读。肽链的合成是从 N 端向 C 端延伸。

由于头部蛋白 80%是由新合成的蛋白质组成。因此他们将各种突变型及野生型 T4 噬菌体侵染 *E. coli* 后 10 分钟，把 ^{14}C 标记的氨基酸加到培养基中，过一段时间，从感染的 *E. coli* 中抽提蛋白，头部蛋白可以通过 ^{14}C 标记来加以鉴别。

他们的实验方法不是对各种突变型的产物测序，而是先将野生型的头部蛋白用胰蛋白酶和糜蛋白酶来处理，消化后所产生的极复杂的混合物中，通过电泳能分离、鉴定出 8 个各有特征的头部蛋白片段（图 14－6），分别是 Cys，T7C(His)，C12b(Tyr)，T6(Trp)，T2a(Pro)，T2(Trp)，C2(Tyr)和 C5(His)片段。然后再测出各 T4 头部蛋白突变型产物含有几个以上的肽段来排序。图 14－7 表示排序的结果和精细作图的序列相一致，不仅表明了基因和蛋白质的共线性关系，同时证明突变型头部蛋白基因内有无义突变的存在，其位置应在各种突变产物的末端。

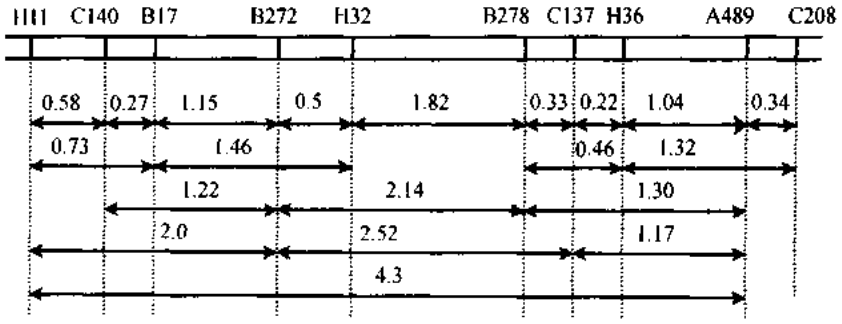


图 14－6 T4 噬菌体头部蛋白基因 23 各种 am 突变的精细结构遗传图。
（仿 Watson, J. D. *et al*: Molecular Biology of the Gene, 4th ed., 1987, Fig. 15－4）

直到 1965 年，Weigert, M. 和 Ggaren, A 由碱性磷酸酶基因中色氨酸位点的氨基酸的置换证明 E. coli 中无义密码子的碱基组成揭示了琥珀和赭石（ochre）突变基因分别是终止密码子 UAG 和 UAA。当时 64 个密码中的 61 个已破译，只留下了 UAA、UAG 和

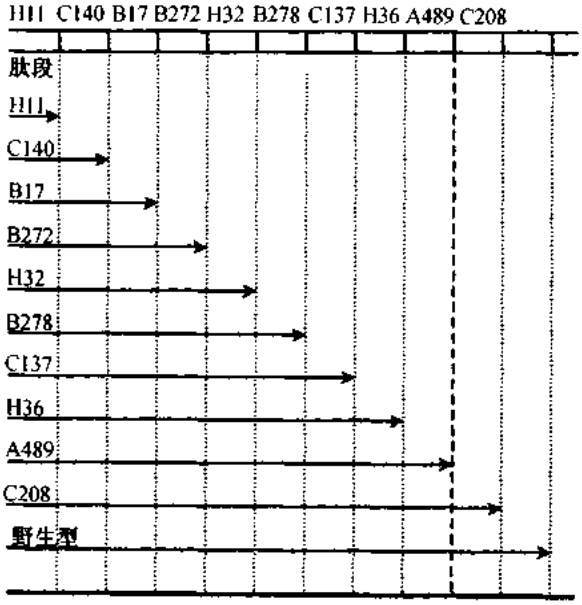


图 14－7 T4 噬菌体头部蛋白基因 23 中的终止密码子。
（仿 P. F. Smith-Kear: Genetics Strauctureand Function, 1975, Fig. 12－10）

UGA 有待确定。Garen 等为了鉴定无义密码子采用了和 Brenner 相似的策略。他们从 *E. coli* 的碱性磷酸酯酶基因（*pho A*）中的一个无义突变品系中分离了大量的回复突变株，然后来探察每一个无义突变中在多肽中相当于已回复的无义密码子位置上的氨基酸究竟是什么氨基酸。结果从图 14－8 中可以看出无义密码子是从该基因的色氨酸位点的密码子产生的。在回复突变中，无义密码子变成了 Trp、Ser、Tyr、Leu、Glu、Gln 和 Lys 的相应密码子。仅有 Trp 的 UGG 变成 UAG，然后在此基础上回复突变成 7 种氨基酸，因此 Trp 产生的无义突变的密码子就是 UAG。最后 1967 年 Brennr 和 Crick 证明 UGA 是第三个无义密码子。根据无义突变的三种昵称，三个终止密码子 UAA 叫赭石（ochre）密

码子(相应于赭石突变)；UAG 叫**琥珀密码子**(相应于琥珀突变)；UGA 叫**蛋白石**(opal)**密码子**(相应于蛋白石突变)。

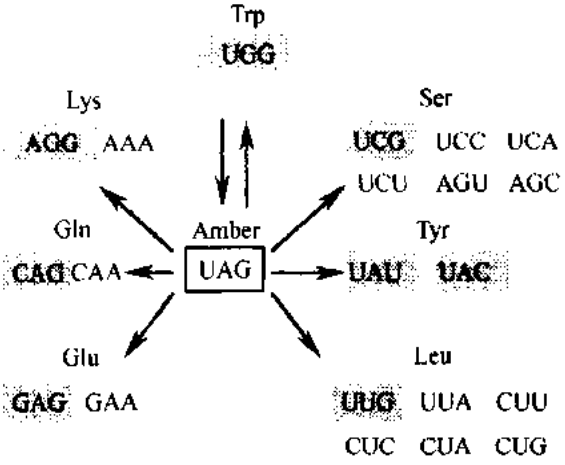


图 14－8　琥珀(am)突变形成的无义密码子鉴定为 UAG。
(Weigert, M. G. *et al.*: Nature, 1965, 206: 992—994)

八、起始密码子的确定

在 Nirenberg 的三联体结合实验和 Khorana 的重复共聚物的体外翻译实验中。合成能从任何碱基起始。但在体内却并非如此，而是需要一个**起始密码子**。什么是起始密码子呢？将各种蛋白质的氨基酸序列和其编码序列相比较，起始时都是 AUG 密码子，读码也都相同，在原核细胞中它编码甲酰甲硫氨酸，在真核细胞中编码未经修饰的甲硫氨酸。

以后又发现在离体的条件下用规则的共聚物为信使时，除了 AUG 外，GUG 也能作为起始密码子，GUG 是 Val 的密码子，由于"摆动"的缘故有时也能和甲酰甲硫氨酸－tRNA 相结合。但摆动的位点并不象一般的情况位于第三位，而是在第一位，使人感到疑惑不解。这可能是由于甲酰甲硫氨酸－tRNA 的反密码子 3′端相邻接的是一个未修饰的腺嘌呤，而在其他 tRNA 中此位点都是烷化的衍生物(图 14－14)。

当初认为以上情况只在离体的条件下才发生，现已搞清 GUC 是 RNA 噬菌体 MS2 的"A"蛋白的起始密码子。而且当正常的 AUG 起始密码子缺失时，GUG 也担当起始密码子的作用。但在离体条件下 GUC 的起始翻译的效率要比 AUG 低得多，可能因为它和甲酰甲硫氨酸－tRNA 的亲和力较低，这也可以作为调控该基因表达的一种手段。

第二节　遗传密码的证实和特点

一、遗传密码的证实

Nirenberg 和 Khorana 用不同的方法破译了 61 个密码子，一年以后(1966 年)俄勒冈大学的 Sterisinger 及其同事用噬菌体 T4 证实了遗传密码是完全正确的。他们采用的方

法跟 Crick 的原黄素诱发移码突变的方法相同，使 T4 溶菌酶产生了移码突变，根据突变后的蛋白质一级结构和野生型溶菌酶氨基酸序列进行了比较，发现仅有 5 个连续的氨基酸不同，其两侧的氨基酸序列都是相同的。他们用遗传密码的字典来拼读这些氨基酸序列，结果鉴定出 Ser 密码子中 A 的缺失，以及 Asn 和 Ala 两个密码子之间插入一个 G(图 14－9)，由于双突变的结果，使读码移框，产生新的密码子正好和突变型的氨基酸序列相吻合。这样不仅证实了遗传密码是正确的，而且根据移码还可以推测出遗传信息是以 5′→3′方向阅读的，蛋白质则是从 N 端向 C 端合成的。

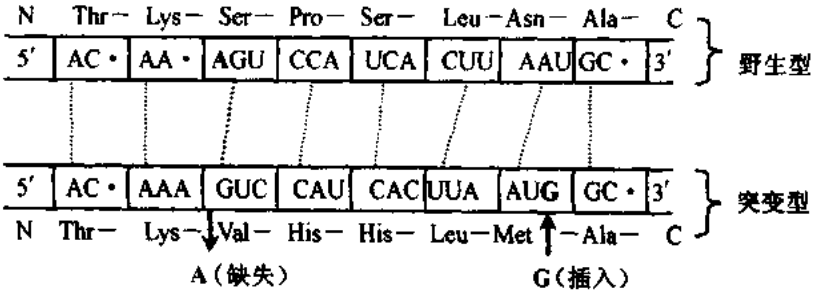


图 14－9　T4 溶菌酶基因经原黄素诱发的插入突变和缺失突变。

二、遗传密码在纤毛虫和线粒体中的改变

遗传密码的普遍性是很严格的，但也存在一些例外的情况。这些变化倾向涉及到起始密码子和终止密码子。或是原来的有义密码子变成了终止密码子，或是原来的终止密码子变成了起始密码子。在表 14－3 中就可以看到在细菌和核基因组中的遗传密码发生改变的 3 种例子。

在原核的枝原体 (*Mycoplasma capricolum*) 中 UGA 并不用于终止，而是 Trp 的密码子。而 UGU 也是 Trp 的密码子，但用得很少。两种 Trp-tRNA 存在，其反密码子是 $\overline{\text{UCA}}$ (阅读 UGA 和 UGG)和 $\overleftarrow{\text{CCA}}$(仅阅读 UGG)。

有的纤毛虫(单细胞的原生动物)将常用的终止密码子 UAA 和 UAG 读成 Glu 的密码子。

表 14－3　密码子在原核生物和真核生物线粒体及原生动物中的改变

生物	AUA	AUG	UGA	UGU	UAA	UAG	CUA	CAA	UAG	AGA	AGG	AAA
一般生物	Ile	Met(起始)	终止	Trp	终止	终止	Arg	Gln	Gln	Arg	Arg	Lys
支原体			Trp	Trp								
纤毛虫					Glu	Glu						
四膜虫					Glu	Glu		Glu	Glu			
游仆虫			Cys			—						
哺乳动物(线粒体)			Trp							终止	终止	Asp
果蝇(线粒体)										Ser		Asp
酵母(线粒体)							Thr					

四膜虫(也属于一种纤毛虫)含有 3 种 $tRNA^{Glu}$。一种是将 CAA，CAG 读成 Glu 一种是将 UAA 和 UAG 读成 Glu(此是符合摆动假说的)，还有一种仅将 UAG 阅读为 Glu，根

据和其他真核生物的比较，进一步改变可能和释放因子 eRF(release factor)有关。

在另一些纤毛虫游仆虫(*Euplotes octacarinatus*)中，UGA 编码 Cys，仅 UAA 用于终止，UAG 未发现。UGA 阅读的改变可能是由于反密码子 $tRNA^{Cys}$ 经过修饰后可以阅读 UGA，而一般是阅读 UGU 和 UGC 的。纤毛虫是从真核生物进化早期分化出来的，人们推测当时的密码仍然有很大的错误倾向，带有不稳定的终止。固定终止的缺乏使得在任何可能的位点上蛋白合成都可以终止。进化的主要分枝中，终止仅在三个密码子处变得有效，使得蛋白质的合成增加了。但在枝原体和纤毛虫中某些终止密码子被保留下来。

三、遗传密码的特点

遗传密码有以下特点：

(1)遗传密码是**三联体密码**。每一个密码子是由 3 个核苷酸构成，它特异地编码多肽链中的一个氨基酸。

(2)遗传密码无逗号。即密码子之间没有空格，阅读 mRNA 时是连续的，一次阅读 3 个核苷酸，也不能跳过任何 mRNA 中的核苷酸。

(3) 遗传密码是不重迭的。阅读 mRNA 时是以密码子为单位，连续阅读。比如 AAGAAGAAG……这一序列，只能释成 Lys-Lys，而不会读成 Lys(AAG) -Arg(AGA) -Glu (GAA)。

(4)遗传密码具有**通用性**。所有的生物都有相同的遗传语言。密码都是相同的。

(5)遗传密码具有**简并性**(degeneracy)，除 AUG(Met)和 UGG(Try)以外，每个氨基酸都有一个以上的密码子，这种现象称为密码的简并(degenecy)。同一个氨基酸的不同密码子称**同义密码子**(synonyms)。表 14－2 表示了简并的特殊模式。当一个氨基酸的密码子头两个核苷酸确定之后第三个可是 U 或 C。比如 Phe 的密码子是 UUU 和 UUC；His 的密码子是 CAU 和 CAC。另一种情况是一个氨基酸的头两个核苷酸确定之后，第三个可以是 A 或 G。如 Leu 的密码子 UUA，UUG；Lys 的密码子是 AAA、AAG；还有一种情况是一个氨基酸的头两个核苷酸确定之后，第三个可以是 U、C、A 或 G。如 Ala 的密码子是 GCU、GCC、GCA 和 GCG。

密码虽有简并性，但它们使用的频率并不相等，有的密码子使用的机会较多，有的几乎不用。如 Leu 有 6 个不同的密码子，但 CUG 使用很高，而 UUA 使用率就很低。每种氨基酸所具有的密码子数目并不和该氨基酸在蛋白质中出现的频率成正比(图 14－10)。

(6)密码子有**起始密码子**和**终止密码子**。蛋白质合成的起始和终止和信号含在密码子中，无论在真核还是在原核中 AUG(Met)都是用作起始密码子，但在少数情况下也用 GUG。

在表 14－2 中还有 3 个终止密码子 (stop codon)，称为**无义密码子** (nonsense codons)或链终止密码子(chain-terminating codons)。它们单个地或串联在一起用于多肽链翻译的结束。当阅读一个特殊的 mRNA 序列时，我们常常要寻找它的起始密码子和终止密码子。终止密码子没有相应的 tRNA 存在，只供释放因子识别来实行翻译的终止，对于一个序列来说是很重要的位点。

(7) 反密码子中的"摆动"(wobble)，由于有 61 个**有义密码子**(sense codons)，那么可能有 61 种 tRNA 具有反密码子分子。但由于反密码子的摆动，对应于整套 61 有义密码子的反密码子明显少于 61 个。原核和真核细胞都只合成约 30 种带有反密码子的

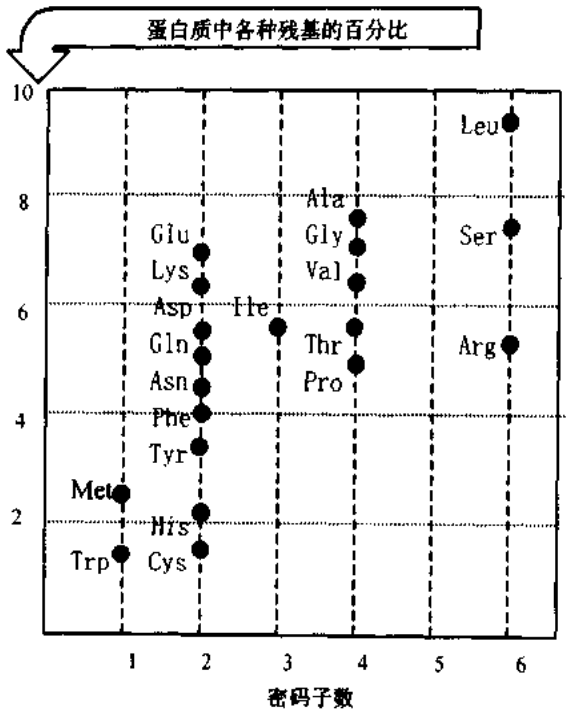


图 14－10　各种氨基酸密码子的数目和氨基酸在蛋白质中出现的频率不成正比。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 9－2）

tRNA。**摆动假说**（wobble hypothesis）是由 Crick. F（1966 年）提出的。即当 tRNA 的反密码子与 mRNA 的密码子配对时前两对严格遵守碱基互补配对法则，但第三对碱基有一定的自由度可以“摆动”（表 14－4，图 14－11）。摆动假说也称为**三中读二**（2 out of 3 reading）。

反密码子第三个碱基的摆动是由于在 tRNA 的三维结构中，反密码子的 5′端的碱基位于“堆积”的最下方（图 14－12），堆积力较小，受到约束也较小，所以可以摆动。

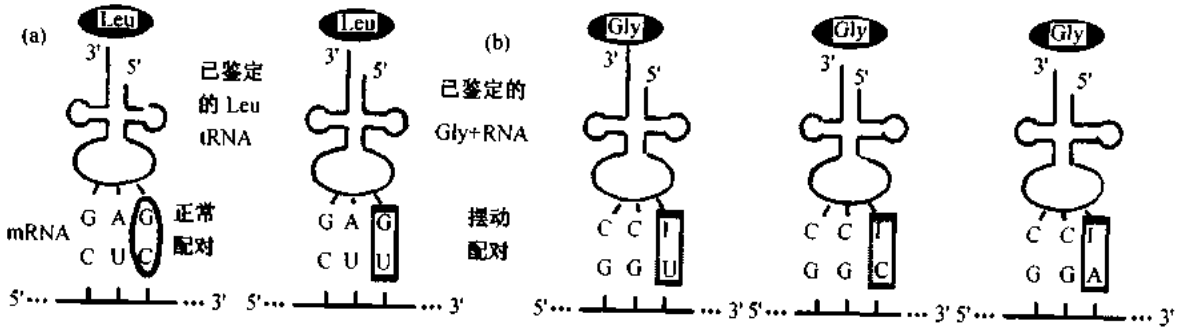


图 14－11　tRNA 反密码子摆动配对。(a) Leu tRNA 的摆动配对，(b) Gly tRNA 的摆动配对。
（仿 P. J. Russell: Genetics, 3 rd. ed., 1992, Fig. 14－5）

根据和反密码子的识别情况遗传密码本身可分成一些组。图 14－13 表示第三个碱基摆动的模式。在八成员组成的密码子家族中，每个成员的四个密码子意思相同，第三个碱基 U、C、A、G 对氨基酸起不到特异的作用。在七个成员组成的密码子家族的意思相同。第三个碱基都是 Py，含 U 或 C。在五个成员组成的密码子家族中，每个成员的

表 14－4　遗传密码中的摆动

反密码子 5′端碱基	密码子 3′端可配对碱基
G	U 或 C
C	G
A	U
U	A 或 G
I(次黄嘌呤)	A、U 或 C

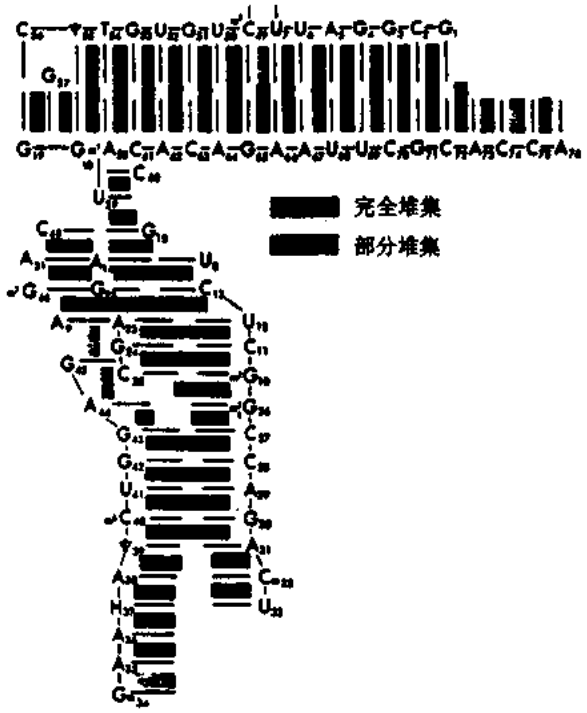


图 14－12　酵母苯丙氨酸 tRNA 分子的碱基堆集作用。深色表示完全堆集，浅色表示部分堆集。细线表示核糖磷酸链，粗线表示嘌呤或嘧啶碱基。第 16, 17, 20 和 47 碱基不参与堆集，图中将它们略去。(转引自 Watson, J.D. *et al*.: Molecular Biology of the Gene, 4th ed., 1987, Fig. 15－4)

二个密码子都是相同的，第三碱基都是 Pu，含 A 或 G 。由一个成员组成的密码子家族中，有 3 个密码子的意思相同，第三个碱基含有 U、C 和 A。

仅有三个密码子的第三个碱基是有作用的，这三个密码子是 AUG(Met)，UGG(Trp)和 UGA(无义密码子)。因此第三位的 C 和 U 总是没有独特的意思，A 也不表明特异的氨基酸。在反密码子和密码子的互补时常常是一个 tRNA 能识别一个以上的密码。这意味着是反密码子的第一位点的碱基和其相应密码子第三位点的碱基的配对必须存在灵活性。即不限于 G：C，A：U 这样的配对(表 14－4)。

三中读二也是有一定规律可循的，一般可分为三种情况：

(1) 密码子的第 1、2 两个碱基和反密码子能形成 6 个氢键时，可三中读二。如 CCX，CGX，GCX 和 GGX。

(2) 密码子的第 1 ，2 两个碱基和反密码子仅能形成 4 个氢键时，不可三中读二。如 AAX，AUX，UAC 和 UUX。

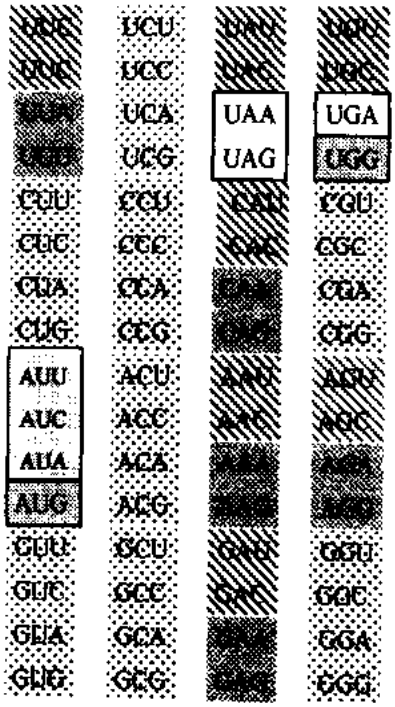
(3) 密码子的第 1，2 两个碱基和反密码子形成 5 个氢键时又分两种情况：当第二个碱基为嘧啶时，可三中读二；如 UCX，ACX，CUX 和 GUX。当第二个碱基为嘌呤时则不能三中读二，如 CAX，GAX，UGX 和 AGC。

以上可以看出能否三中读二和 tRNA 与密码子结合的稳定性有关。

第三节　tRNA 的结构和功能

在介绍 tRNA 的结构和功能之前让我们来回顾一下一段有趣的历史，它反映了科学家们的睿智和远见卓识。

早在 1955 年 Crick F.H.C 在考虑 DNA 与蛋白质合成之间的相关性时就指出，有些**转接器**(adaptor)分子可以把相应的氨基酸携带到 DNA 链的适当位置。转接器或适配器这个名词是由 Brenner 提出的。以上是 Crick F.H.C 的"变性模板与转接器"的假说。当时并不知道 tRNA 的存在，更不知道其结构。那么组成多肽的氨基酸如何在模板的指导下合成蛋白质呢，Crick 就设想应由一种小分子物质－转接器来负责特异氨基酸的运载。



第三个碱基和密码子的关系	同义密码子的第三个碱基	密码子数
与第三个碱基无关	U,C,A,G	32
Py 和 Pu 之差	U或C	14
	A或G	10
仅限一种情况	U,A,C	3
	仅G	2

图 14－13　第三个碱基和密码子的关系。(仿 B.Lewin:Genes Ⅵ,1997,Fig.9－3)

1956 年 Volkin，E 和 Astrachan，L 发现噬菌体 T2 侵染 *E.coli* 后，出现了和噬菌体碱基组分相应的 RNA，从而获得初步的－mRNA 的认识。同年 Hoagland,M.B 等先后在真核和原核细胞中发现了一类小分子 RNA，长 80 个 nt，含量稳定，但又不是核糖体的结构成份。这些短链的 RNA 分子终于被定名为**转运 RNA** 或 **tRNA**。P.Zamenic 等发现了在蛋白质生物合成中氨基酸被利用后 tRNA 参与蛋白质的合成才被人所共知。

1957 年 9 月 Crick 在一次会上作了"蛋白质合成"的报告，1958 年公开发表了这篇报告，提出转接器分子含有核苷酸使得氨基酸序列的正确确定以及复制和转录的精确性取决于碱基互补。他写道："我首先天真地想到 RNA 可能是由形成 20 个不同空穴的构型构成的，其每个"空穴"适合于 20 种氨基酸中的一种……，一个自然而然的假设是氨基酸被一个转接器分子携带到模板上去。而转接器有一部分能配合到 RNA 上"。Crick 的设想已将翻译和 tRNA 之间的关系给人们勾画了出来。

一、tRNA 的结构

1. 三叶草型的二维结构

tRNA 的结构是在 1965 年由 Holley，R.N. 及其同事对酵母丙氨酸 tRNA 的研究首先发现的，因此他荣获了 1968 年度的诺贝尔化学奖。几百种来自细菌和真核生物的 tRNA 的序列通过碱基的互补配对都能形成**三叶草型**(cloverleaf)的结构。图 14－14 表示其结构，其特点是：

(1)各种 tRNA 均含有 70～80 个碱基，其中 22 个碱基是恒定的。

(2)其 5′端总是配对的，这和 tRNA 的稳定性有关。5′端和 3′端配对(常为 7bp)形成茎区，称为**受体臂**(acceptor arm)或称**氨基酸臂**。在 3′端永远是 4 个碱基(XCCA)的单链区，在其末端有 2′－OH 或 3′－OH，是被氨基酰化位点。此臂负责携带特异的氨基酸。

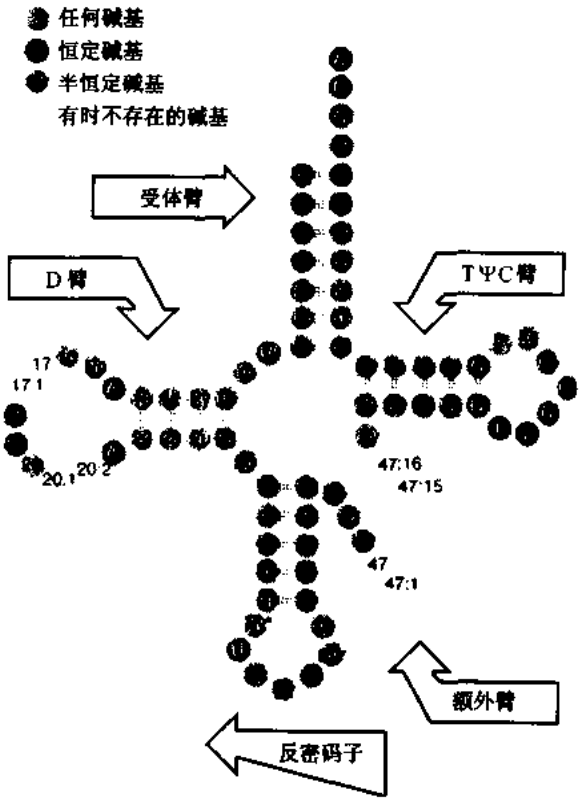


图 14－14　tRNA 的三叶草型结构。

其他的臂都为茎环结构。

(3)**TΨC 臂**存在着特殊的碱基 Ψ(假尿嘧啶)，常由 5bp 的茎和 7nt 和环组成。此臂负责和核糖体上的 rRNA 识别结合；

(4)**反密码子臂**(anticodon arm)常由 5bp 的茎区和 7nt 的环区组成，在环区的中央总存在着反密码子三联体，它负责 mRNA 上的密码子的识别与配对。

(5)**D 环**(D arm)的茎区长度常为 4bp，环区有 4 个碱基不恒定，此环在 17，17：1

20：1,20：2 位置含有特殊的碱基 D(双氢尿嘧啶),故也称为(双氢尿嘧啶环)。

(6)**额外环**(extra arm)可变性大,从 4 nt 到 21 nt 不等,其功能是在 tRNA 的 L 型三维结构中负责连接两个区域(D 环 - 反密码子环和 TΨC - 受体臂)。

2. tRNA 的三维结构

tRNA 的三叶草结构并没有使我们明确 tRNA 的各个臂在空间上如何排列,只有通过 X 衍射来研究 tRNA 的晶体才能获得 tRNA 真实的构象。从几种酵母的 tRNA 晶体的

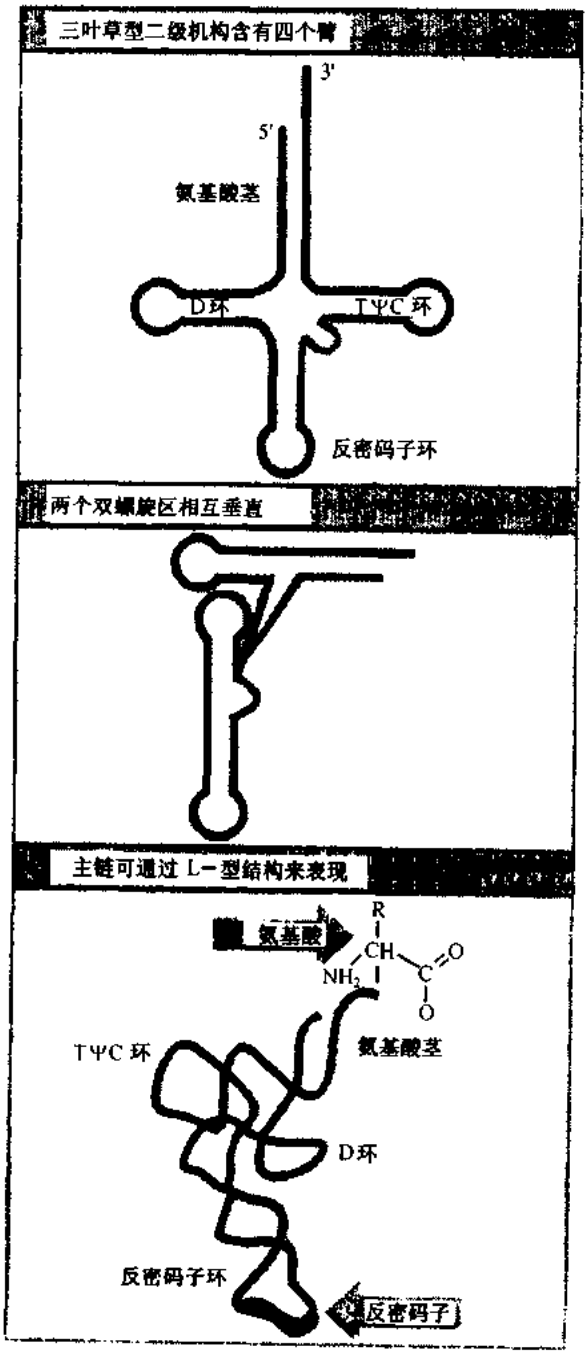


图 14 - 15 tRNA 由三叶草型折叠成 L 型三维结构。氨基酸茎在一端,反密码子在另一端。
(引自 B.Lewin: Genes Ⅵ,1997,Fig.7 - 5)

研究表明，各种 tRNA 存在共同的三维结构。它主要是依靠茎中的碱基配对来维持，形成了稳定的 **L 型结构**(图 14－15)。此结构可描绘如下：

(1)氨基酸受体臂位于 L 型的一侧，距反密码子环约 70Å。

(2)D 环和 TΨC 环形成了"L"的转角。

(3)在一些保守和半保守的碱基之间形成很多的三级氢键(图 14－16a)，使分子形成 L 形(图 14－16b)，并使结构稳定。

(4)使得三维结构得以形成的这些碱基配对不同于 DNA 双链中的碱基配对。涉及到与磷酸核糖主链相互作用的三级结构的磷酸二酯键分布在核糖的 2′－OH 上。

(5) 几乎所有的碱基其朝向使得碱基平面之间产生堆积的作用。这种作用也是 tRNA 构象稳定的主要因素之一。

(6)在反密码子茎中仅有很少的三级氢键，这可能由于在蛋白质合成时，反密码子区的相对方向会发生改变之故。

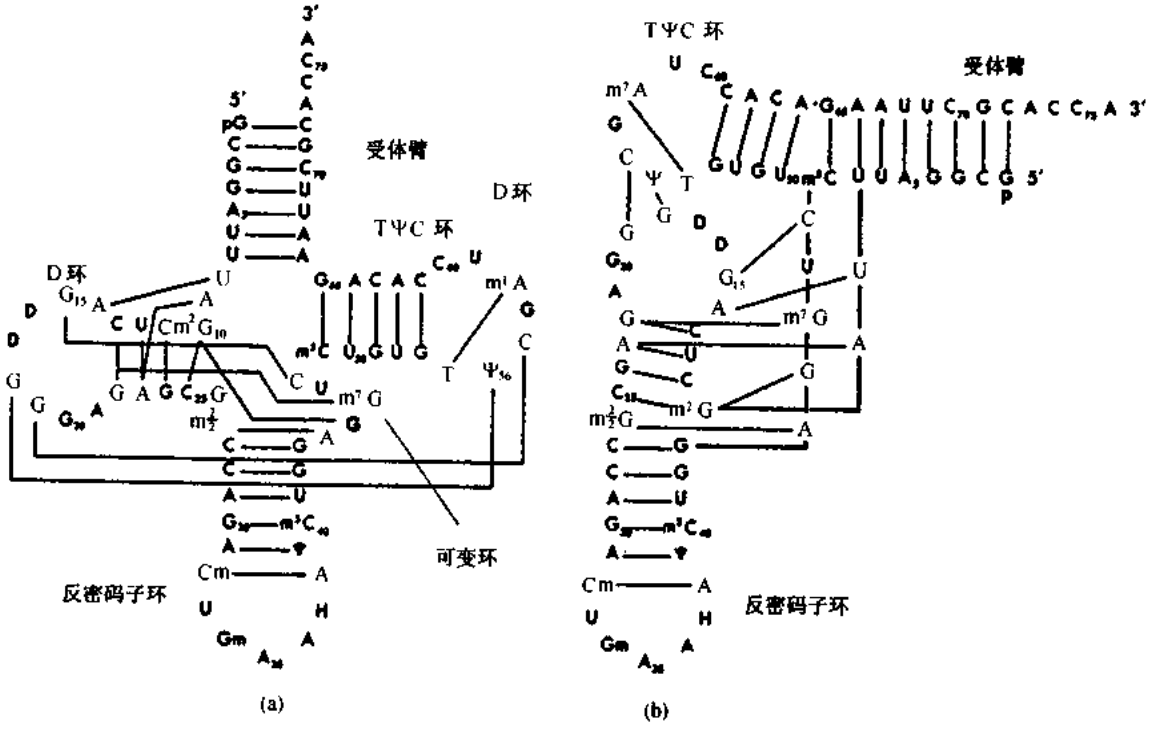


图 14－16　酵母苯丙氨酸 tRNA 的三维结构。(a)三叶草模型；(b)L 型三维结构。方框表示可变的碱基，深色细线表示三级氢键。(转引自 Watson, J. D. *et al*.: Molecular Biology of the Gene, 4th ed., 1987, Fig. 15－4)

二、tRNA 对氨基酸的识别

tRNA 通过反密码子和 mRNA 上的密码子相互配对，将特定的氨基酸运送到核糖体上肽链合成位点上，但是 tRNA 如何来识别特定的氨基酸呢？这就涉及 tRNA 的"身份"(identity)问题，这个问题是核酸领域的热点之一。人们需要解决几个问题：① tRNA 怎样接受特定的氨基酸，氨基酰－tRNA 合成酶怎样识别 tRNA；②tRNA 中的哪些结构和接受特定氨基酸有关。

1988 年 Hou Ya-ming(侯雅明)和 Schimmel 首先取得了这方面的突破。他们采用的

方法是：选用发生色氨酸琥珀突变(UGG→UAG)的营养缺陷型 *E.coli* (trp^-)来进行研究；这种突变型必须在加有色氨酸的培养基上才能生长。携带着 Ala，反密码子突变成 CUA 的校正 tRMA，可以和终止密码子 UAG 相配对，可以校正色氨酸的琥珀突变，在色氨酸突变位点加入了 Ala 。用含有校正 tRNA 的基因转化 *E.coli* 后只要校正率达到 3%，该 trp^- 菌株就可在普通培养基上生长。然后他们用点突变的方法来改变校正 tRNA(Ala)上的各个位点，观察对识别 Ala 有何影响，以此来确定与识别特异氨基酸的有关位点。他们获得了 28 个突变株。分别转化 *E.coli*(trp^-)，再检测转化后的 *E.coli*(trp^-)在基本培养基上是否能生长。若仍能生长，说明相应的点突变位点和氨基酸的识别无关，若不能生长说明相应的点突变位点与 tRNA 的"身份"有关。

他们用上述方法证明了 Ala tRNA 的 G3：U70 碱基对决定了丙氨酰 tRNA 合成酶与 tRNA 的识别，他们就把这样很小的元件称之为 tRNA 的"identity"，或称为**副密码子**(paracodon)。到目前为止已知的 tRNA 的"身份"如表 14－5 所示。

表 14－5 每种合成酶通过几个特殊碱基来识别其同质 tRNA

tRNA	合成酶识别的碱基
一类氨基酰 tRNA 合成酶	
Val	反密码子上的三个碱基
Met	反密码子上的三个碱基
Ile	反密码子上的 C34 修饰碱基
Gln	U35(反密码子)；U1－A72 和 G73(受体臂)
二类氨基酰 tRNA 合成酶	
Phe(酵母)	反密码子上的三个碱基，G20(D 环)；A73(末端)
Ser	G1－C72；G2－C71；A3－U70(受体臂)；C11－G24(D 环)
Ala	G3－U70(受体臂)

Schimmel ， Hou ya-ming 和施建平等用合成的微螺旋做氨基酸接受试验，证明只要具有 Ala tRNA 受体臂那样长的微螺旋，或者具有该 tRNA 受体臂和 TΨC 茎连在一起的小螺旋，且在受体臂有相当于 G3－U70 的碱基对，它们就能接受 Ala，已发现带有了三种不同反密码子(CUA，GGC，UGC)的 Ala tRNA 都具有 G3－U70 副密码子。

表 14－5 表明这些位置的共同特点是：①关键位点的数目很少(1～5)个；②副密码子较集中在反密码子和受体臂上，这和 tRNA 的 L 型三维结构和氨基酰－tRNA 聚合酶结合的空间构型有关(图 14－16)；③由于一个碱基对不能足以使 20 套 tRNA 被普遍的识别，因此不同的 tRNA 都有自己特殊的规律来进行识别。

三、校正 tRNA

分离 tRNA 突变体的最有效的方法是分析 tRNA 对 mRNA 中的密码子所具有的反应能力，并确定 tRNA 分子不同部位对密码子反密码子识别所起的作用。

有些 tRNA 的突变体能抑制基因突变，正如前面我们已经描述过的基因回复突变中就已涉及到**抑制基因**(suppressor)或称**校正基因**。

在 tRNA 校正系统中，第一次突变是发生在 mRNA 上，改变了某个密码子，使有义密码子变成了无义密码子，因而蛋白质的产物是截短的或平截的，一般都是没有功能的。而第二次发生的抑制突变或校正突变是改变 tRNA 上的反密码子，这样它仍可识

别取代了原来靶密码子的突变密码子，插入了一个氨基酸，使蛋白质的功能得到恢复。此抑制称为依赖于原来突变特点的无义(nonsense)或错义(missense)抑制。

在一个野生型的细胞中，无义突变只被一种释放因子所识别而终止蛋白质的合成。**校正突变**(suppressor mutation)产生了一种氨基酰－tRNA 能识别终止密码子，通过一个氨基酸的插入使蛋白质的合成能持续下去(图 14－17)。

翻译系统这种新的能力使完整的蛋白质得以合成。如果抑制 tRNA 所携带的氨基酸和野生型蛋白此位点原有的氨基酸不同的话，此蛋白质的活性可能会降低。

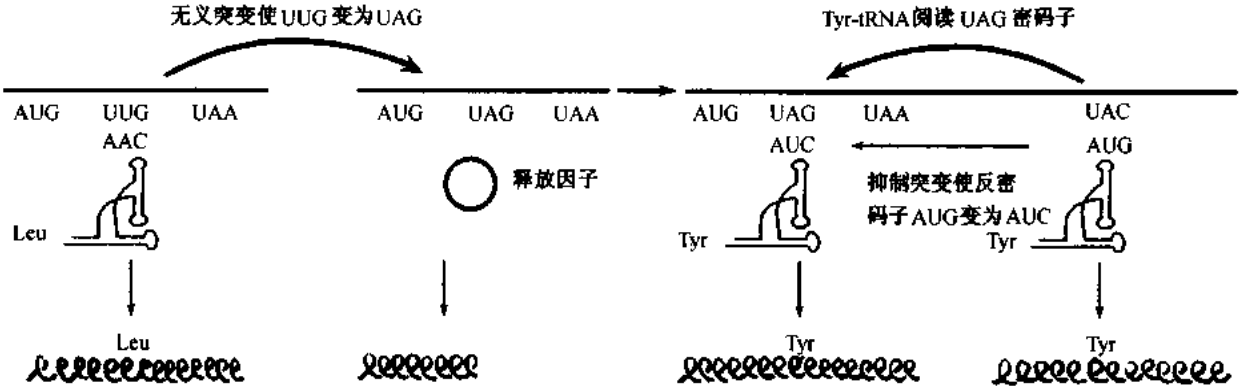


图 14－17　带有突变反密码子的 tRNA 可抑制无义突变。

1. 无义抑制

通过抑制 tRNA 识别无义突变位点，将某种氨基酸插入该位点，使得多肽链继续延伸，而不中途停止，这种抑制就称为**无义抑制**(nonsense suppressor)或无义校正。

无义抑制是通过三个不同的途径来进行的：①tRNA 反义密码子的突变；②副密码子或其他结构的改变；③tRNA 反密码子化学修饰。

(1)tRNA 反密码子的突变

此途径是抑制基因的主要途径。根据抑制基因的反密码子，无义抑制分为三类，分别对应于三个终止密码子。表 14－6 表示几种研究得最多的抑制基因。

最容易研究的是琥珀突变(am)抑制基因。在大肠杆菌中至少有六种 tRNA 突变后

表 14－6　由反密码子突变而产生的无义抑制基因

基因	tRNA	野生型		抑制基因	
		识别的密码子	反密码子	反密码子	识别的密码子
SupD(*su1*)	Ser	UCG	CGA	CUA	UAG
SupE(*su2*)	Gln	CAG	CUG	CUA	UAG
SupF(*su3*)	Tyr	UAC，UAU	GUA	CUA	UAG
SupC(*su4*)	Tyr	UAC/UAU	GUA	UUA	UAA/UAG
SupG(*su5*)	Lys	AAA/AAG	UUU	UUA	UAA/UAG
SupU(*su7*)	Trp	UGG	CCA	UCA	UGA/UGG

B. Lewin: Genes V, 1994, Table 8－5

可识别密码子 UAG，所有的 am 抑制 tRNA 都有反密码子 CUA，它是由各种野生型 tRNA 反密码子中单个碱基的改变而产生的(表 14－6)。

(2)tRNA 其他结构的改变

另一种类型的 UGA 抑制基因有一个意外的特点，它来自相同的 $tRNA^{Trp}$，但它仅仅是在 tRNA D 茎的 24 位点上由 A 取代了 G，使 G·U 对变成 A·U 时，从而增加了双链的稳定性。其反密码子的顺序仍保持和野生型的相同，为 CCA。那么 D 茎上的突变必然以某种方式改变反密码子环的构象，使 CCA 上的 C 能和 UGA 上的 A 进行非同寻常的配对，这种抑制 tRNA 仍能识别通常的密码子 UGG。在真核的 tRNA 中也存在相关的反应。牛肝中含有一种带有反密码子 $m\overbrace{CCA}$ 的 $tRNA^{Ser}$。按照摆动的法则，这种 tRNA 将对 Trp 的密码子 UGG 作出反应，但实际上它识别终止密码子 UGA，在这种情况下 UGA 可能自然地被抑制。

以上的发现表明无论是野生型的密码子和反密码子的识别，还是突变型的反密码子和密码子的识别都不能完全凭借三联体密码子的顺序，还会受到分子其他特点的影响。

(3) tRNA 上反密码子的化学修饰

所有的无义抑制基因都是通过编码的 tRNA 的结构基因上的点突变而产生的。从原则上说，抑制 tRNA 也可能通过其反密码子上碱基的化学修饰的变化而产生。这种变化是编码 tRNA－修饰酶的基因突变所产生的。

2. 错义抑制

错义突变使得其编码的氨基酸变成另一种氨基酸，这种新的氨基酸取代了蛋白质中某位点上原来的残基可能使蛋白质失去功能（任何氨基酸的取代构成的错义突变，可以通过蛋白质活性的改变而被检测）。这种突变能通过插入原来的氨基酸或其他的氨基酸而被抑制。

图 14－18 表明错义抑制基因能通过无义抑制相同的方法来起作用。即 $tRNA^{Gly}$ 反

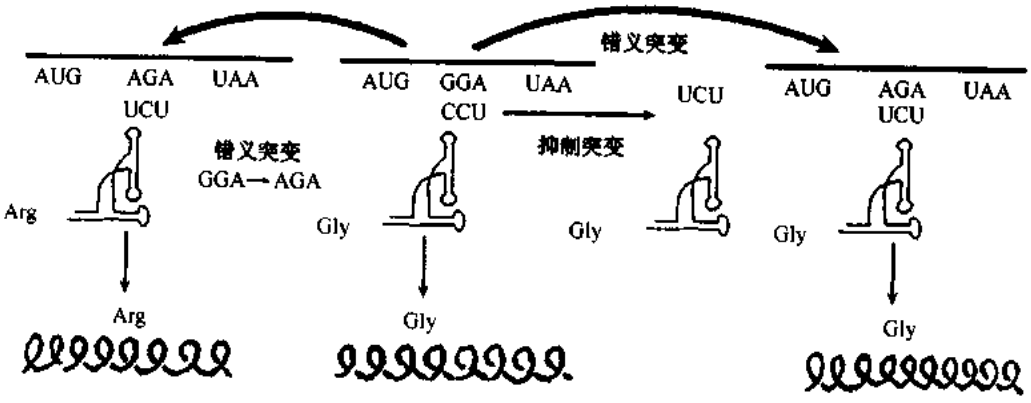


图 14－18 反密码子发生突变可抑制错义突变。

密码子发生突变，可以对已突变的密码子作出反应。错义抑制涉及到将原来的氨基酸改为另外的氨基酸。只要能完全恢复或部分恢复蛋白质活性，都称作为**错义抑制**。

另一种途径也可产生错义抑制。即抑制发生在 tRNA(或氨基酰 tRNA 合成酶上)对氨基酸的识别与连接的位点上。虽然这种 tRNA 的反密码子没有改变，但它所携带的氨基

酸却不同了，这也可能导致错义抑制。

抑制突变的特点：

(1) 不是所有终止密码子的抑制基因都能产生有功能的蛋白质，起到抑制或校正的作用，关键要看氨基酸取代的情况。比如由 Leu 的密码子 UUG 突变成 UAG 时，当抑制基因所携带的氨基酸是 Tyr、Ser 或 Trp 都可能起到抑制的功能，若是氨基酸是带电荷的 Lys、Gln 和 Glu，则蛋白质可能是没有功能的。

(2)校正的作用不可能是完全的。这是由两方面的因素决定的。①校正的 tRNA 分子是有限的而且还要和释放因子竞争，这样使得无义突变位点并不能全部得到校正；②若是错义抑制的话，由于氨基酸发生取代，使得蛋白质的活性有所降低。

(3)反密码子中的三个碱基中的任何一个都可能发生突变。如表 14－6 中的 *supD*，*supE* 和 *supF* 三个抑制基因反密码子的突变位点都不同。每种抑制 tRNA 一般都只识别 UAG 终止密码子，而不再识别原来相应的密码子。

(4) 赭石突变抑制基因不仅可以识别赭石密码子 (UUA)，也可以抑制琥珀突变(Am)码子 UAG。但反过来 Am 抑制基因(CUA)就不能抑制赭石突变(UAA)，这是由于"摆动"缘故所造成。

(5)当细胞中含有多个 tRNA 拷贝时，抑制才能发挥作用。

(6)有的抑制基因，不仅可以识别终止密码子，而且还可以识别原来的密码子。如 UGA 的抑制基因，有一种类型(supU)是 $tRNA^{Trp}$ 的突变体。野生型 $tRNA^{Trp}$ 的反密码子是$\overleftarrow{CCA}$，它是识别密码子 UGG 的。抑制 tRNA 具有突变的反密码子$\overleftarrow{UCA}$，它不仅可以识别原来的密码子 UGG，而且还可以识别终止密码子 UGA。

(7) 校正基因一般不会影响正常的终止，但有时也会使正常的终止抑制而造成肽链的增长。这是由于：①校正基因识别的终止密码子不一定和正常终止的密码子相同。有时正常终止位点有两个连续的终止密码子，而且结构不同，如 UAG-UAA；②释放因子将和抑制基因竞争和终止密码子的结合；③抑制基因的效率很低，通常为 1% ~ 5%，所以常不会抑制正常终止。

抑制基因不仅对无义突变作出反应，同时对正常的终止密码有时也能抑制，使正常的多肽链发生了延长，其功能可能也会发生改变，因此 UAA 抑制基因的存在使细胞生长比正常的要慢得多。由于 UAA 终止密码子作用强，使用的频率比较高，因此另外两个无义密码子 UAG 和 UGA 的抑制基因的影响要比 Ochre 抑制来得小。而且 Ochre 抑制基因不仅抑制 UAA，还可以识别 UAG 终止密码子，所以其影响之大远比 Amber 和 Opal 抑制要明显得多。

第四节　核糖体的结构和功能

核糖体就像一个小的可移动的工厂，沿着 mRNA 这一模板，不断向前迅速合成肽链。氨基酰 tRNA 以极高的速率进入核糖体，将氨基酸转到肽链上，有从另外的位置被排出核糖体，延伸因子也不断地和核糖体结合和解离。核糖体和附加因子一道为蛋白质合成的每一步骤提供了活性区域。

细菌的核糖体是附着在 mRNA 上，DNA 一边转录成 mRNA，核糖体就一边结合上去，沿着 5′→3′方向进行翻译，形成转录和翻译的偶联。在真核的胞浆中核糖体总是和细胞骨架或粗面内质网的膜相结合。其共同特点是核糖体在细胞中从事翻译时并不是游离的，总是直接或间接地和细胞结构相连接。

所有的核糖体都可分为大亚基和小亚基两部分。每一亚基都含有主要的 rRNA 成分和很多不同的蛋白质分子。大亚基含有较小分子的 RNA，这 rRNA 分子虽然是由许多基因编码的，但在一个细胞中它们的序列却没有什么变化。有人认为这是因为变化序列受到自然选择而被淘汰。

rRNAs 的主要功能是结构。蛋白质和每种主要 rRNA 的特殊位点相结合，以特定的序列装配成亚基。核糖体的有趣之处不仅在于其功能，而且还在于它们本身由 RNA 和蛋白质装配而成的过程。

核糖体具有各种活性中心，每一活性中心都由特定的一组蛋白和 rRNA 的一定区域装配而成。这些活性中心要求结构中的 rRNA 的直接参与或者具有催化作用。有的催化功能需要个别蛋白。但还没有一种活性由于蛋白质的分离或一组蛋白的分离而减少。其实目前并没有证据表明核糖体蛋白直接地涉及任何催化的功能。

核糖体是很多蛋白质的集合体。每一种活性是和各种结构相关的。这些活性共同来完成翻译的作用。很多的蛋白(和 rRNA)是和建立各种结构有关，这些结构使各种活性位点之间存在一种适当的关系。我们应当将核糖体看成为一种相互作用的结构。在此结构中构象的改变将遗传信息从一个位点转移到另一位点。

两种不同类型的信息对于分析核糖体是重要的。突变涉及到参与特殊反应中的特殊核糖体蛋白或 rRNA 的碱基。结构分析，包括核糖体成分的直接修饰和比较 rRNA 的结构，鉴别了与特殊功能有关的各个物理位置。

一、核糖体的组成

细菌和哺乳动物核糖体的组成如表 14－7 所示。现在来比较一下两者的成分。两者都是核糖体蛋白颗粒，它们含 RNA 的量要多于蛋白质。这种核糖体蛋白被称为 **r－**

表 14－7 原核和真核生物核糖体的组成及功能

核糖体	亚基	rRNAs	蛋白	RNA 的特异顺序和功能
细菌 70S	50S	23S = 2904b	34 种(L1－L34)	
2.5×10^6 d		5S = 120b		含 CGAAC 和 GTΨCG 互补
66%RNA	30S	16S = 1541b	21 种(S1－S21)	16S RNA(CCUCCU)和 S-D 顺序(AGGAGG)互补
哺乳动物 80S	60S	28S = 4718b	50 种	
4.2×10^6 d		5.8S = 160b		有 GAUC 和 $tRNA^{fMet}$的 TΨCG 互补
60%RNA		5S = 120b		
	40S	18S = 1874b	33 种	和 Cap m^7G 结合

蛋白(r-protein)。

细菌中的所有的核糖体成分都已被鉴定，在 *E. coli* 中核糖体 rRNAs 都已被测序。其小亚基(30S)是由 16S rRNA 和 21 种蛋白构成的。大亚基(50S)是由 23S rRNA 和小分子的 5S rRNA 以及 34 种蛋白构成的。除一种蛋白的核糖体存在四种拷贝以外，其余的蛋白都只有一个拷贝。

高等真核生物的核糖体要比细菌的大得多。RNA 和蛋白的总量很大。主要的 RNA 分子较长，并含有多种蛋白质，RNA 在量上仍占主导地位。

细胞器的核糖体和胞浆中的核糖体是不同的，具有各种不同的类型。最大的差不多和细菌的核糖体大小相似。含 70% 的 RNA，最小的仅有 60S，所含 RNA 少于 30%。

细菌、叶绿体和真核细胞质核糖体的形态都十分相似（线粒体的核糖体的形态尚不清楚）。图 14－19 是细菌核糖体亚基的大体形态。小亚基形态是平的。而大亚基的是凸形的结构。

完整的 70S 核糖体具有不对称的结构，在小亚基的头部和中部与大亚基结合形成了 V 形的隧道。图 14－20 表示细菌核糖体亚基及完整核糖体的电镜观察结果。

哺乳动物细胞质核糖体的外形是相似的，但比较大。表明细菌的核糖体和真核生物核糖体的小亚基是相似的，而大亚基明显不同。虽然精确的模型尚未建立，但哺乳动物的核糖体的大小亚基之间的关系看来和原核的核糖体是相似的。

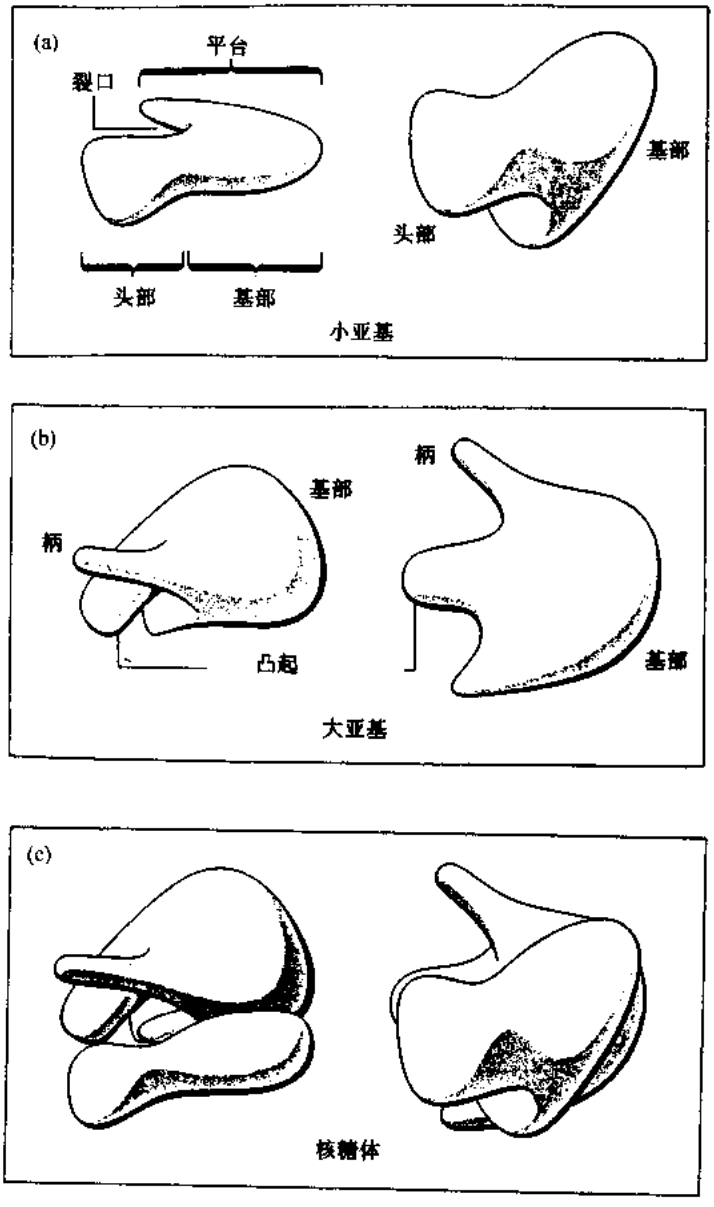


图 14－19　细菌核糖体的形态结构。(a)30S 小亚基；(b)50S 大亚基；(c)由大、小亚基结合装配成的核糖体。（引自 B. Lewin: Genes Ⅵ, 1990, Fig. 9－1、2、3）

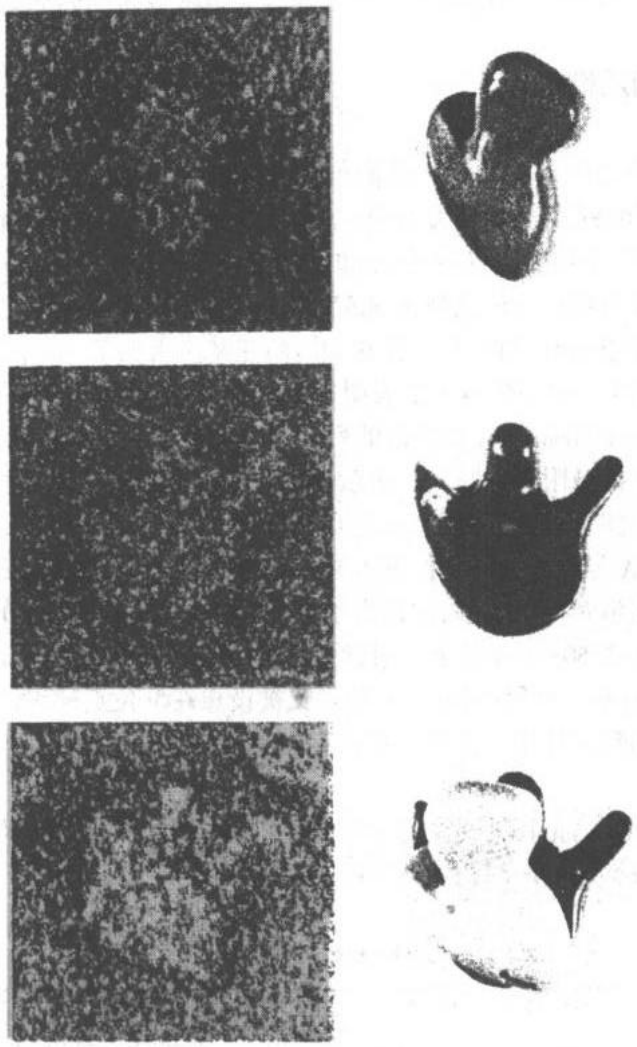


图 14－20　细菌核糖体及其大、小亚基的电镜照片和模型。(引自 B.Lewin: Genes Ⅵ, 1997, Fig. 8－2)

核糖体中 mRNA 的结合位点是人们所关注的。一种普遍的观点认为此位点应在亚基之间或紧靠连接点。有的模型把 mRNA 画成穿过核糖体，但结合位点可能是位于核糖体表面。图 14－21 表明两个 tRNA 的大小与核糖体的比例关系。看来 tRNA 可能插

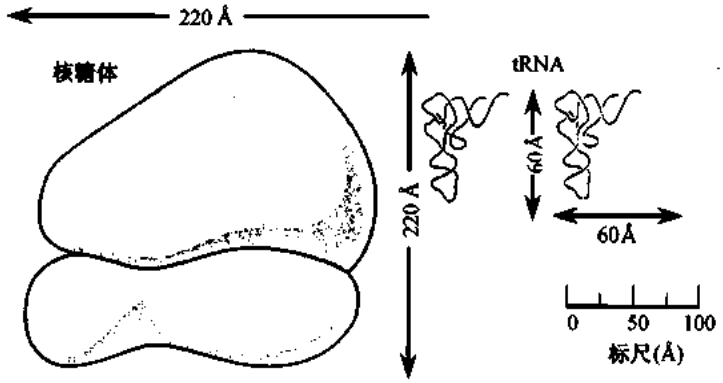


图 14－21　比较核糖体和 tRNA 的大小，表明核糖体足以和 2 个 tRNA 结合。(可结合 35 个碱基的 mRNA)(参考 B.Lewin: Genes Ⅵ, 1997, Fig. 8－5)

入到大的开放表面"裂口"(Clefts),而不会插入完整核糖体的内部空洞中。

二、核糖体的活性位点

根据核糖体的位点和功能，我们可以区分出核糖体的几种活性。30S 亚基结合 mRNA 和起始 - tRNA 起始因子复合体。然后它们结合 50S 亚基。70S 核糖体可能具有供 tRNA 结合的两个功能位点：A 位点和 P 位点。肽转移酶的中心在 50S 亚基上，EF-G 结合位点以及负责移位的位点都在 50S 大亚基上(图 14 - 22)。

结合位点是很大的，每个位点都在相应的亚基上占据了一定的位置。各个功能域就像酶的活性中心一样。图 14 - 22 表明这些位点。它们占据了核糖体的 2/3，称为**翻译功能区**(translational domain)，核糖体的另外的区域是**多肽出口的功能区**(exit domain)，核糖体通过这个区域附着在膜上。多肽链从一个功能区出现进入另一个功能区，通过一个离肽酰转移酶有一定距离的出口伸出来。

tRNA 进入 A 位点通过移位再进入 P 位点。再通过 E 位点离开细菌的核糖体。A 位点和 P 位点必定涉及两个亚基，这是由于 tRNA 要和 30S 小亚基上的 mRNA 配对，但肽链的转移又发生在 50S 大亚基上。通常设想 A:P 两个位点离得很近，这样易位时便于 tRNA 由一个位点移入到另一位点。E 位点又被设想在 P 位点的附近。它和 A 位点之间具有一种变构的相互作用。由于 E 位点存在脱酰基 tRNA，使得 A 位点对氨基酰 tRNA 的亲和力减弱。

这些位点真实的拓扑学关系是怎样的呢？核糖体的成分又涉及哪些功能呢？人们用各种方法来研究它们的结构与功能的关系(见表 14 - 8)。

表 14 - 8　核糖体的活性位点

活性位点	功能	位置	组分
mRNA 结合位点	结合 mRNA 和 IF 因子	30S，P 位点附近	S1、S18、S21；及 S3、S4、S5、S12 16S rRNA3′末端区域
P 位点	结合 fMet-tRNA 和肽基- tRNA	大部分在 50S 亚基	L2、L27 及 L14、L18、L24、L33 16S 和 23S rRNA3′附近区域
A 位点	结合氨酰基- tRNA	大部分在 30S 亚基	L1、L5、L7/L12、L20、L30、L33 16S 和 23S rRNA(16S 的 1400 区)
E 位点	结合脱酰 tRNA	50S	23S rRNA 是重要的
5S rRNA	和 23S rRNA 结合	P 和 A 位点的附近	L5、L18、L25 复合体
肽酰基转移酶	将肽链转移到氨基酰- tRNA 上	50S 的中心突起	L2、L3、L4、L15、L16 23S rRNA 是重要的
EF-Tu 结合位点	氨基酰- tRNA 的进入	30S 外部	
EF-G 结合位点	移位	50S 亚基的界面上，L7/L12 附近，近 S12	
L7/L12	GTP 酶需要	50S 的柄	L7、L12

1. mRNA 的结合位点

30S 亚基和 mRNA 结合的起始需要 S1 蛋白，这种蛋白对单链核苷酸的亲和性很强。它负责维持 mRNA 以单链状态结合在 30S 亚基上。这样可防止 mRNA 自身进行碱

基配对，以免影响翻译的稳定性。

S1 在核糖体中延伸得很长，其位置也十分复杂。经交联(cross-linking)研究表明，它和 S18、S21 密切相关，当起始 tRNA 和 AUG 密码子结合时，这两种蛋白是存在的。S1、S18 和 S21 这三种蛋白构成了涉及 mRNA 起始结合和起始 tRNA 结合的一个结构域。这样它应当位于小亚基裂缝附近的 mRNA－结合位点。其 rRNA 的 3′端在此区域要和 mRNA 起始位点配对。

起始因子结合在核糖体的相同区域。IF3，可能会涉及到 mRNA 的结合。如果小亚基的 IF3 结合区域涉及到和大亚基的结合，那么 IF3 的作用可能是使 mRNA-30S 小亚基稳定结合，然后它再被 50S 大亚基所置换。

现在还不清楚亚基连接的机制。当 16S rRNA 环上(第 791 位点)的碱基发生突变时会影响到亚基的连接。而此位点正好位于两个亚基的界面上。在此处还有 23S rRNA 和 16S rRNA 3′区域互补的部分。有可能这些碱基配对和亚基的彼此结合有关。IF3 和 16SrRNA 结合与两亚基的联接是不相容的。

16S rRNA 是属于 A 位点的部分，此 rRNA 在 3′端小结构域中有一段保守序列，长约 16Nt(1392～1407)，称为 **1400 区**。此区可能和 tRNA 相互作用。交叉－连接研究表明肽酰 tRNA 可与此区域连接。当氨基酰 tRNA 结合到核糖体上时，这一区域的一些碱基受到保护而不和化学探针结合。

有关 EF-Tu 的交连研究表明，A 位点位于一组蛋白的附近，但目前还不太清楚这些相互作用的蛋白怎样为 tRNA 提供结合位点。与此相同在 P 位点的肽酰 tRNA 3′端周围的一组蛋白已被鉴定。

组建成 50S 亚基的 5S rRNA 在体外也能依赖三种蛋白 L5、L8 和 L25 装配成复合体。此复合能和 23S rRNA 结合，现在尚未能分离复合体的成分。5S rRNA 位点位于 P 位点和 A 位点的附近。

肽基转移酶位点位于由**嘌呤霉素**(puromycin)结合的中心突起处。一组蛋白和 23S rRNA 涉及到此位点的产生。

增长的多肽链在离肽基转移酶位点约 150Å 处从 50S 亚基中挤出。肽的延伸在通过核糖体时是以一种非折叠肽链的形式存在的，当离开出口功能区，呈游离状态时就开始折叠。

除核糖体蛋白以外，还存在 L7/L12。L7 和 L12 不同，仅在 N－末端存在乙酰基。在每个核糖体上有两个拷贝的二聚体(L7/L12)，L7/L12 组装形成了大亚基的柄(stalk)。当把它们除去时，蛋白就不能应任何附加因子的要求来执行 GTP 的水解。这并不是说 L7/L12 就是 GTPase，它可能对于另一些蛋白的激活是所需要的。

虽然翻译要涉及到 mRNA 在 30S 亚基上的移动，但易位因子 EF－G 是结合在 50S 亚基上的，EF－G 的结合位点紧靠 S12，S12 是 30S 亚基上 mRNA 结合位点的蛋白之一。这样 EF－G 的位置是在亚基的界面上，就在 L7/L12 二聚体的附近。

通过 23S rRNA 的改变而得知 EF－G 和 EF－Tu 位点是相连接的。EF－G 和核糖体结合是在 23S rRNA 的两个位点上，一个在第 2660 位点的发夹环上，这个位置同时也受到 EF-Tu 的保护。在真核 28S rRNA 的这一相邻位点是抑制延伸的外源凝集素的靶位点，显然这一"环"的结构对于延伸来说是很重要的。

P 位点和 A 位点必须靠得很近。tRNA 用反密码子结合在 mRNA 上。P 位点影响到 A 位点的激活，这是因为肽酰基－tRNA 必须存在并与氨基酰 tRNA 相结合，核糖体必须保持同时容纳两个相邻的 tRNA 分子，相邻的两个反密码子之间的距离不能大于

10Å 左右。tRNA 的直径大约是 20Å。

tRNA 的构象看来在 A、P 位点要保持相同。需要解决的一个问题是 mRNA 的两个密码子之间可能发生扭曲或结节，若如此的话，两个相邻的 tRNA 与密码子结合之后方向可能不同。核糖体移动时就会产生与 tRNA 从 A 位点移到 P 位点的方向不一致。mRNA 必须具有促进 tRNA 移动的功能，这样可以用反密码子和密码子的配对来度量易位的距离。

各种核糖体活性位点之间在功能上的相互关系现在尚没有完全弄清楚。核糖体可能有一种高度相互作用的结构，在这些结构中某一位点发生改变就会明显地影响到其他位点的活性。

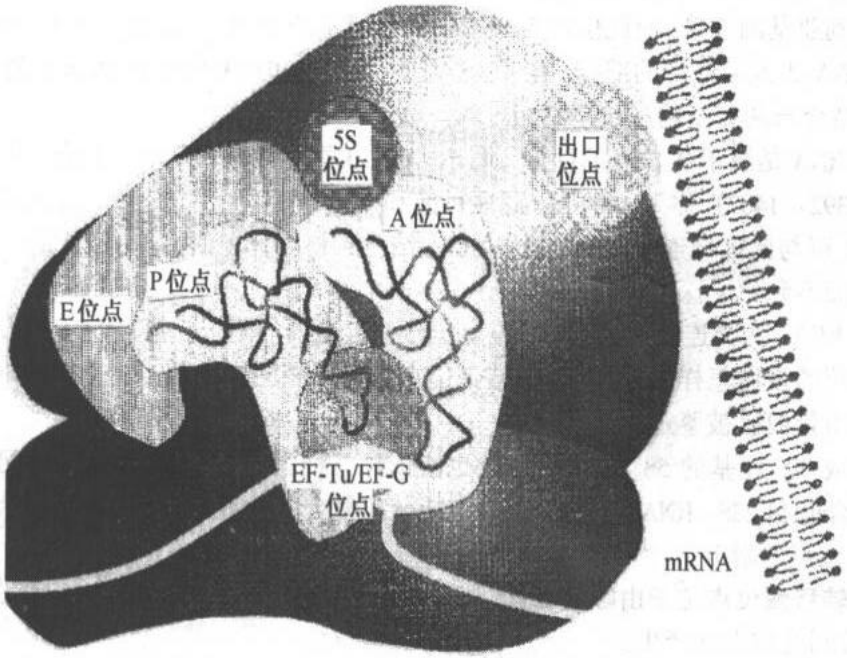


图 14－22　*E.coli* 核糖体的几个活性中心。(仿 B.Lewin: Genes Ⅵ, 1997, Fig.8－27)

第十五章

蛋白质的合成，转运和加工

基本概念

氨基酸进入蛋白质合成途径是通过氨基酰－tRNA 合成酶，这种酶将氨基酸和特异的 tRNA 连接起来。氨基酰合成酶通过构象、反应动力学及化学等不同途径分别对氨基酸和 tRNA 进行鉴别校对。

细菌核糖体有 A、P、E 三个位点，A 位点为受体位点（acceptor site）。P 位点为供体位点（donor site），被肽酰－tRNA（peptidyl-tRNA）占据。E 位点是核糖体上脱酰－tRNA 结合位点，E 和 A 位点互为反协同效应。

起始因子 IF3 和 30S 小亚基结合使其保持游离状态，30S 和 50S 亚基结合前 IF3 要释放出来。起始 $tRNA^{f met}$ 首先进入 P 位，其结构和 $tRNA^{Met}$ 不同。

IF2 和 $tRNA^{f met}$ 结合形成二元复合体。在 30S－mRNA 复合体上形成完整的起始复合体。

真核生物肽链合成的起始原核生物复杂得多，涉及到各种起始因子如 CBP，4A，4B，eIF-Ⅰ，eIF-Ⅱ等。

原核 16S 的 3′端顺序和 S-D 顺序互补，对翻译的起始起着重要的作用。

原核中 Tu 和 Tg 是延伸因子，Tu 使 aa-tRNA 进入 A 位，Tg 是一个循环因子，使 Tu 从 Tu-GDP 转换成 Tu-GTP。

肽链的合成是由肽转移酶催化的。

肽链合成的终止必须有释放因子的介入，它们识别终止密码。原核细胞有 3 种终止因子 RF-1，RF-2，RF-3；真核细胞只有一种释放因子。

真核细胞中蛋白质的转运有两种类型：翻译后转运和协同转运。

翻译后转运是由导肽将蛋白导向核、线粒体、叶绿体及其他细胞器。

协同翻译转运是由信号肽将蛋白导向 ER，继而运向高尔基体、溶菌体、泡囊、质膜及分泌途径。

导肽的功能主要和其结构有关而和顺序无关，顺序也不保守。识别信号是二级结构，常由疏水氨基酸和碱性氨基酸组成，越膜主要是依赖双亲螺旋的形成，膜电位的存在以及表面受体的识别。

导肽在细胞器中的定位依赖第二个信号——靶信号。

信号肽由疏水核心构成。

SRP 是一种 RNP，可以和信号肽结合，制动肽的延伸及和膜上的 SRP 受体结合。

信号肽的锚顺序可以使蛋白跨膜，并决定蛋白在膜上的朝向。

细菌的越膜运输也有翻译后转运和协同翻译转运两种。前者采用孔径途径分泌到膜外，后者采用普遍性途径越过内膜。

普遍性途径需要分子伴侣、外围膜蛋白 SecA、膜本体蛋白 SecE、SecY、导肽酶及 SRP 同源蛋白的参与。此途径仅可越过内膜，外膜仍不易越过。

孔径途径需要分泌蛋白、内膜蛋白和跨膜蛋白的参与。

翻译后加工主要包括 N 端 f-Met 和 Met 的切除，二硫键的形成，化学修饰和前体的剪切。

化学修饰的主要内容是糖基化。

新生肽在 ER 中的折叠和修饰密切相关。

在 ER 中开始进行初步糖基化，多萜醇是寡糖的载体，依赖糖基转移酶将寡糖加到蛋白的 Asn 上（N－糖苷键），在糖苷键的作用下对寡糖进行剪切。在高尔基体中进一步完成修饰。

蛋白质的剪切分为三类，一是将多蛋白切开成为几个成熟蛋白，二是切除中间部分，余下的部分由二硫键连接，三是蛋白内含子被切除，外显子重新连接在一起。

第一节　氨基酰 -tRNA 的形成及核糖体的作用位点

前面我们已介绍了 tRNA 的结构和功能，核糖体的结构和活性位点以及遗传密码，现在介绍遗传信使翻译的细节，即蛋白质合成的过程。蛋白质的合成是在核糖体上进行的，编码遗传信息的 mRNA 在核糖体上被翻译。翻译的方向是从 mRNA 的 5′端到 3′端，和转录的方向一致。多肽合成的方向是从 N 末端(氨基端)到 C 末端(羧基端)。氨基酸结合在相应的 tRNA 分子上，被转运到核糖体上。肽链上氨基酸顺序的排列是依赖下列因素决定的：①mRNA 上的密码子和 tRNA 上的反密码子的互补结合；②每种特异的氨基酸能和与其相应的 tRNA 结合。

mRNA 的密码子识别 tRNA 的反密码子，而不识别 tRNA 所带的氨基酸，这是由 G.von Ehrenstein，B.Weisbllum 和 S.Benzer 提出的(1962)。他们以一个十分漂亮的实验证实，在体外一旦半胱氨酸连接到正确的 tRNA 上，它可能被化学修饰成丙氨酸，但这并不影响它们和肽链的连接，结果使组合到肽链中的丙氨酸代替了半胱氨酸。Ala－$tRNA^{Cys}$ 被用于体外合成血红蛋白。在体内珠蛋白的 α 链和 β 链都含有一个 Cys。当进行珠蛋白体外合成时，α 和 β 链上原来是 Cys 的位点结果出现了 Ala，这就意味着丙氨酸－$tRNA^{Cys}$ 是可以阅读 Cys 密码子并插入到肽链中。这表明特异密码子的阅读依赖于 tRNA 分子而不依赖于它所携带的氨基酸。

蛋白质合成分为 3 个阶段：起始，延伸和终止。这在原核和真核生物中都是相似的，下面我们将详细地介绍每一阶段的情况，原核的蛋白质合成多以 *E.coli* 为例的，原核生物和真核生物在翻译中的重要差别是应当注意的。

一、氨基酰 tRNA 分子的形成

tRNA 分子在蛋白质合成中的重要功能是将特异的氨基酸运送到核糖体 mRNA 的复合体上，那么氨基酸如何和 tRNA 结合呢？首先要经过氨基酸的活化，然后再形成氨基酰－tRNA。

氨基酸进入蛋白质合成途径是通过氨基酰－tRNA 合成酶，这种酶将氨基酸和特异的 tRNA 连接起来。所有这种合成酶的功能都是分两步来进行的，其机制如图 15－1 所示。

首先是氨基酸和 ATP 反应形成氨基酰－腺苷酸，氨基酸＋ATP＋E→氨基酰－ATP－E＋PPi　释放出焦磷酸，能量是由剪切 ATP 的高能磷酸键所提供的。

然后活化的氨基酸转移到 tRNA，释放出 AMP。以上以反应式表示：

活化：氨基酸＋ATP＋E→氨基酰－ATP－E＋PPi

转移：氨基酰－ATP－E　＋tRNA→aa－tRNA＋AMP＋E

这种合成酶将 tRNA 和氨基酸分检到相应的位点，每种合成酶都能识别单一的氨基酸和所有的对应 tRNA，然后将相应的氨基酸和 tRNA 装配起来。合成酶对 tRNA 的识别描述成它的**关连或同族 tRNA**(cognate tRNA)。通常它们的反密码子是不同的。

这种合成酶尽管功能相同，但可分成不同的类型。他们的分子量从 40kDa 到 100kDa 大小不等，有的是单体，而有的是双体或三体。它们之间的同源性很小，可能是在进化的早期产生的，保留了与 tRNA 及氨基酸相互作用这一功能。当然识别

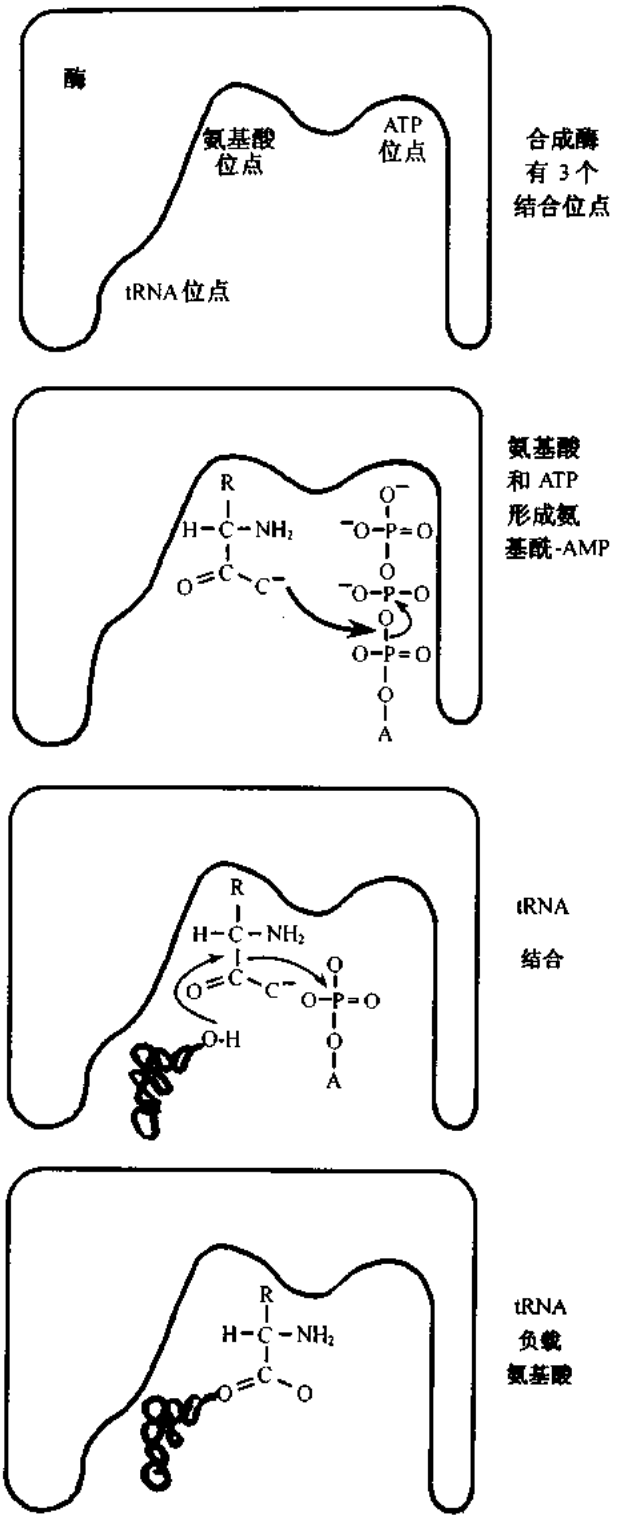


图 15－1　氨基酰合成酶催化产生 aa-tRNA。(B.Lewin: Genes Ⅵ, 1997, Fig.9－7)

tRNA 的活性位点是由分子的小部分构成。将不同合成酶的活性位点进行比较可得到有趣的结果。

根据其 tRNA 结合区的结构氨基酰合成酶一般分为两组，每组含有 10 种酶（图 15－2)。催化功能域包括 ATP 和氨基酸结合位点，它可能作为一个大的功能域被识别，其间插入了 tRNA 受体螺旋结合功能域。这种排列方式使 tRNA 的末端接近催化位点。另一个分离的功能区结合 tRNA 的反密码子区。只有多体(multimeric)合成酶才具

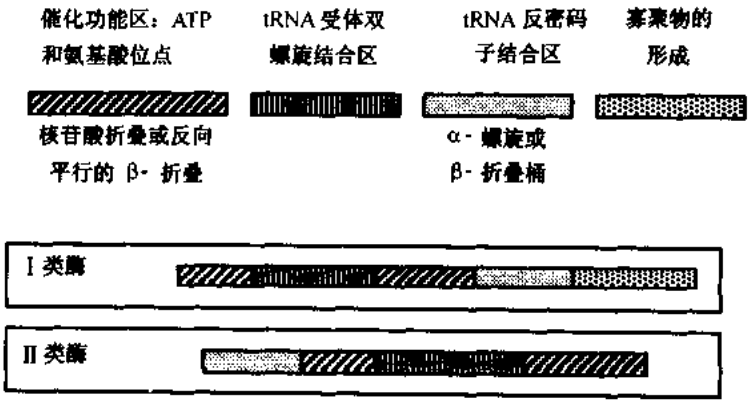


图 15－2　氨基酰 tRNA 合成酶含有 3－4 个不同的功能区。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 9－8)

有寡聚物形成功能域(oligomerization domain)。

Ⅰ类酶具有 N 端催化功能域，它存在两个短的氨基酸保守顺序，有时称其为“标识顺序”(signature sequence)。此催化功能域是以称为**核苷酸结合折叠**(nucleotide binding fold)的一种基序的形式而存在(这种基序在其他与核苷酸结合的酶中也有发现)。此核苷酸折叠是由交替平行排列的 β－链和 α－螺旋所构成。识别顺序构成了 ATP－结合位点的部分。在不同的Ⅰ类酶中间，插入在催化功能区的 tRNA 受体螺旋也是不同的。Ⅰ类酶的 C 端功能域包括 tRNA 反密码子结合功能区及寡聚物形成功能域，不同的Ⅰ类酶这个区域也是有差别的。

Ⅱ类酶在它们的催化功能域中由三种相似的顺序构成。活性位点含有一个大的反向排列的 β－折叠(β－Sheet)，周围围绕着 α－螺旋。插入在催化功能域中的 tRNA 受体螺旋功能域有一个结构，在不同的Ⅱ类酶中是不同的。反密码子结合区位于 N 端。而寡聚物形成功能域的位点广泛存在差异。

在两组合成酶之间缺乏明显的相关性，这是一个谜。有可能它们彼此独自进化。有一种可能是它们早于生命的形成，而以蛋白质的形式存在。

合成酶·tRNA 结合的一般模型表明蛋白质和 tRNA 的结合是沿着 L 型分子的侧面结合的。同样的原则适用于所有合成酶·tRNA 结合：tRNA 的结合原则上在两个末端，大部分 tRNA 顺序都不涉及被合成酶识别的问题。但 tRNA 和合成酶相互作用的特点在Ⅰ类酶和Ⅱ类酶之间是不同的。我们从图 15－3,4 所示根据结晶所作的模型中就可看到这一点。这两种类型的酶和 tRNA 不同的侧面相接近。从 tRNA－蛋白质的模型来看，彼此呈镜像对称的关系(图 15－3,4)。

Ⅰ类酶(Gln－tRNA 合成酶)接近 tRNA 的 D 环一侧，而在受体茎末端的结合位点识别其小沟，同时也识别另一端的反密码子环。此结构所显示的特点是 tRNA 和酶接

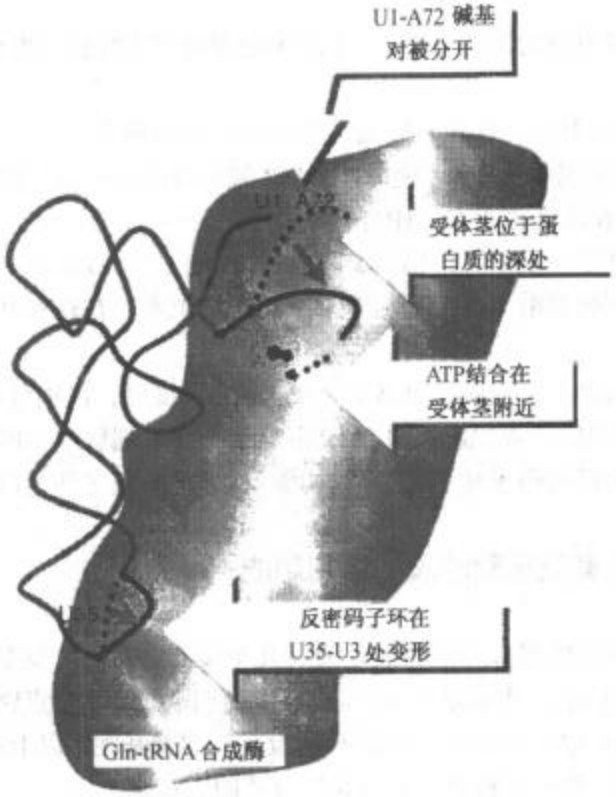


图 15－3　Ⅰ类 tRNA 合成酶与 tRNA 受体茎的小沟和反密码子环结合。
（引自 B.Lewin:Genes Ⅵ,1997,Fig.9－10）

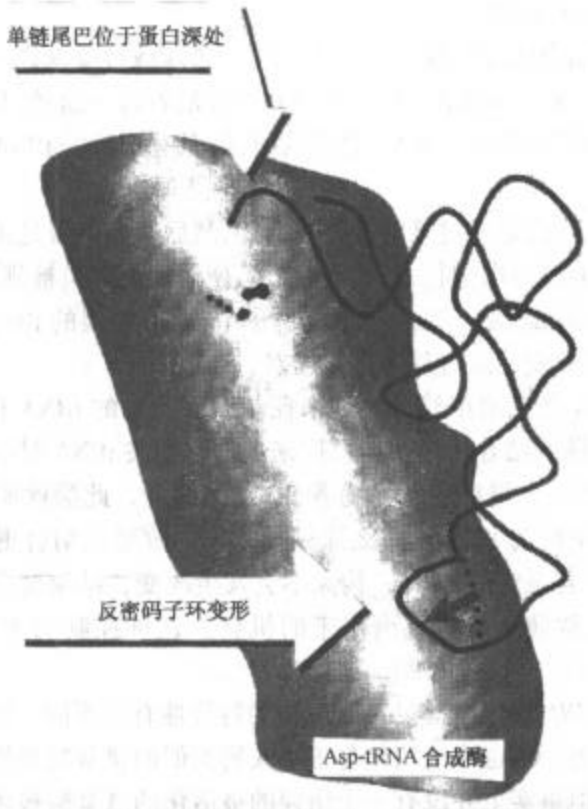


图 15－4　Ⅰ类 tRNA 合成酶，在 tRNA 受体茎的大沟和反密码子环处与其结合。
（引自 B.Lewin:Genes Ⅵ,1997,Fig.9－10）

触后改变两个重要位点的结构，比较一下图中的虚线和实线就可看出反密码子环和受体茎的变形。

(1)反密码子环上的 U35 和 U36 被拉开而进入蛋白质中；

(2)受体茎的末端严重变形，使 U1 和 A72 碱基对被分开，此茎单链末端插入合成酶蛋白质的深处，在此处也含有 ATP 结合位点。

这个结构印证了早期研究的结果，即当 tRNA 的 U35，G73 或 U1 与 A72 碱基对若发生改变会影响到氨基酰 - tRNA 合成酶对 tRNA 的识别。在这些 tRNA 和蛋白质之间都可产生氢键。

Ⅱ类酶（Asp-tRNA 合成酶）是从另一侧与 tRNA 接触，识别可变环及受体茎的大沟。受体茎保持着规则的双螺旋构象，ATP 可能结合在末端腺嘌呤的附近。在结合位点的另一端紧密地和反密码子环结合，环的构象发生改变，随之和蛋白质紧密结合。

二、合成酶在负载氨基酸阶段的鉴别功能

在基因表达中鉴别氨基酸的特点是通过几个步骤进行的。氨基酰 - tRNA 合成酶是如何识别相应的 tRNA 和氨基酸呢？核糖体如何识别与 A 位点密码子相应的 tRNA 呢？在合成 DNA 或 RNA 时酶如何识别和模板互补的碱基呢？以上存在着一个共同的问题，如何从一套底物中鉴别出一个特异的合适的成员来。

可能任何一个成员都能通过一个随机碰撞的过程和活性中心相结合，但接着错误的成员被排斥，只有正确的成员才被接纳。正确的成员总是占少数（20 个氨基酸中的一个，约 40 个 tRNA 中的一个，4 个碱基之一），因此标准应当是严格的。我们可以从排斥和接纳这两种结果来推测。

(1)在单一的结合阶段可能存在进入，审查，排斥或接受这样一个循环。下一阶段再涉及是否反应。也就是说结合位点的亲和性控制着这一途径，只有特别适合的才能留在结合位点。对于氨基酰 - tRNA 合成酶而言，只有正确的 tRNA 才能在结合位点稳定的结合。

(2)选择，这个反应是通过某些途径进行的，然后确定是否是正确的 tRNA 存在。若不存在，这个反应向反方向进行或者产生绕道，使错误的成员被排斥。这种结合后审查一般称为校对（proofreading）。在合成酶的例子中，若有错误的 tRNA 或氨基酸存在，则可能通过中心阶段改变反应过程来进行校对。

氨基酰 - tRNA 合成酶用校对机制来控制识别正确的 tRNA 和氨基酸。图 15 - 5 表明 tRNA 和合成酶的结合通过两步反应来进行，相关 tRNA 对结合位点有很高的亲和性，因此结合较快，解离较慢。随着 tRNA 的结合，此酶就要“审查”这个被结合的 tRNA。若是正确的 tRNA，那么通过酶构象的改变使结合更为稳定。接着迅速的发生氨基酰化。若是错的 tRNA，构象不会发生改变，结果反应过程变得很慢，这样增加 tRNA 在负载前从酶中解离出来的机会。这种控制的类型称为**动力学校对**（kinetic proofreading）。

各种氨基酰 tRNA 合成酶对氨基酸底物的特异性有所不同。有些对一种氨基酸是高度特异结合的，另一些也可能活化与正确底物类似的氨基酸虽然氨基酸的类似物有时也能被腺苷化，但事实上并没有一个错误的被活化的氨基酸被用于形成一个稳定的氨基酰 - tRNA。

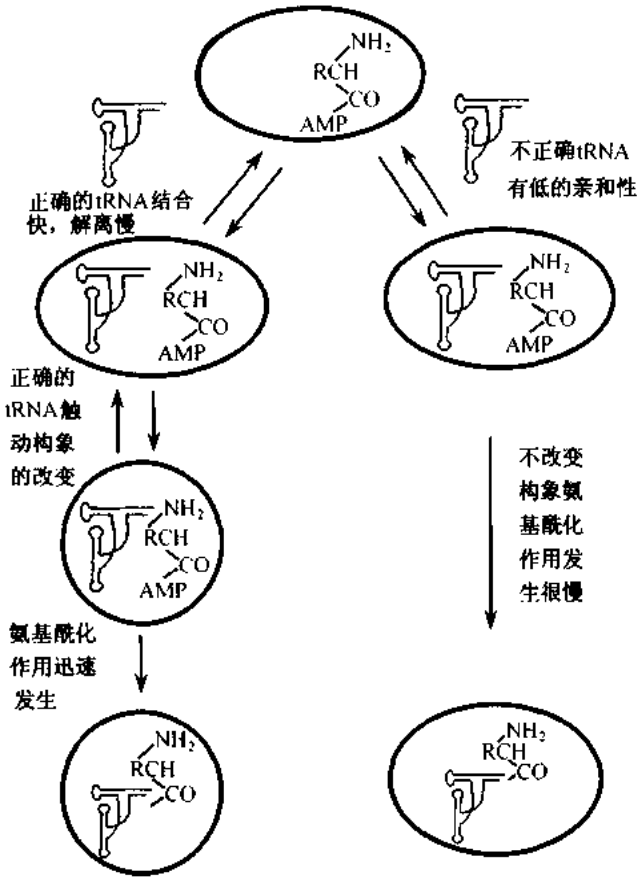


图 15－5　氨基酰－tRNA 合成酶通过 2 个步骤控制对正确 tRNA 的识别。(1) 酶对正确的 tRNA 有很高的亲和力；(2)不正确的 tRNA 发生氨基酰化很慢。(参考 B.Lewin：Genes Ⅵ，1997，Fig.9－12)

依赖于反应的部分途径然后再通过逆反应进行校对称为**化学校对**（chemical proofreading)。在校对一个错误氨基酸时有两个阶段可以进行。一个是校对错误的氨基酰－腺苷（图 15－6)。另一个是校对错误的氨基酰－tRNA。两个步骤都需要正确的 tRNA 存在。对于整个的合成酶来说两个途径中有一个占优势，至于哪一种途径占优势也是可以变化的。

在正确的 tRNA 结合时错误的氨基酰－腺苷可被水解。此机制在几种合成酶中是占优势的，包括 Met，Ile 和 Val(通常这种反应在体内不可能观察到，当错误的氨基酸是同型半胱氨酸(它缺乏甲硫氨酸的甲基)时，可能出现这种反应。校对的适当形式释放出错误的氨基酸。如作为同型半胱氨酸内脂。事实上同型半胱氨酸内脂(homocysteine thiolactone)在 *E.coli* 中是作为 Met-tRNA 合成酶负载反应的副产物产生的。这表明校对是 tRNA 负载氨基酸过程的一部分。

后一步骤化学校对的在有些合成酶中占优势。错误的氨基酸实际上转到了 tRNA 上，然后在 tRNA 结合位点被识别，因而被水解并释放出来。这一过程需要一个连接和水解的持续的循环，直到一个正确的氨基酸转移到 tRNA 上为止。若校对发生在氨基酰－腺苷形成之前的阶段，正确的 tRNA 的存在对于触动校对是必要的。一个经典的例子是在大肠杆菌中 Ile-tRNA 合成酶依赖 tRNA 的存在来区别氨基酸。这个酶能负载带有 AMP 的 Val，但当添加 $tRNA^{Ile}$ 时，Val 酰－腺苷被水解。总的错误的比率都是取决

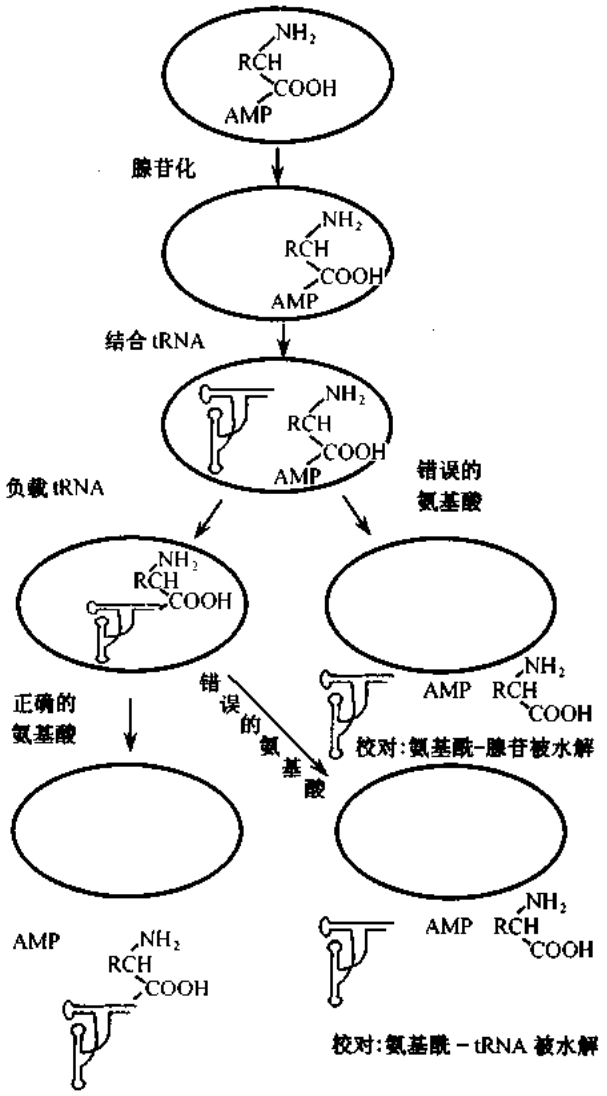


图 15－6　氨基酰 tRNA 合成酶和错误的氨基酸结合时，校对需要正确 tRNA 的结合，它或使酶构成发生改变，导致错误的氨基酰－腺苷的水解，或使氨基酸转移到 tRNA，随之错误的氨基酰－tRNA 被水解。（参考 B. Lewin：Genes Ⅵ，1997，Fig. 9－13）

于个别阶段的错误频率。如表 15－1 所示。

表 15－1　负载 $tRNA^{Ile}$ 的精确性通过它的合成酶对二个阶段错误率的控制

步骤	错误频率
Val 对 Val-AMP 的激活	1/225
Val-$tRNA^{Ile}$ 释放	1/271
总的错误的比率	1/255 × 1/271 = 1/60,000

B. Lewin: Genes Ⅵ, 1997, Table 9.2

三、核糖体的作用位点

蛋白质的合成涉及到一个组装线，在组装线中核糖体固定不变地延着 mRNA 前进，引进氨基酰 tRNA，为蛋白质合成提供氨基酸。核糖体本身作为一种因子，它由蛋白质和 rRNA 紧密结合形成各种活性中心，承担着各种催化活性。在蛋白质合成的三个阶段中，不同的附加因子来帮助核糖体合成的起始、延伸和终止。核糖体活动所需要的能量由水解 GTP 来提供。

(1)起始(initiation)涉及在蛋白质头两个氨基酸之间肽键形成的反应。它需要核糖体结合 mRNA，形成含有第一个氨基酰－tRNA 的起始复合物。这是在蛋白质合成中相对较慢的一个步骤，通常它决定 mRNA 翻译的速率。

(2) 延伸 (elongation) 包括从合成第一个肽链到最后一个氨基酸的加入所有的反应。氨基酸每次加入一个。这是在蛋白质合成中氨基酸加入最快的步骤。

(3)终止(termination)此步骤包括完整肽链的释放，同时核糖体从 mRNA 上解离来。

虽然翻译的速度在很大的程度上取决于温度，但蛋白质的合成整个过程都是很快的。在细菌中，37℃时每秒可在肽链上增加约 15 个氨基酸，因此平均在约 20 秒的就可以合成 300 个氨基酸的肽。在真核生物中蛋白质合成的速率要慢一些，在卵细胞里，37℃时每秒肽链只加上约 2 个氨基酸。

大部分用于了解蛋白质合成的实验都是借助于体外系统来完成的，体外系统含有核糖体，氨基酰－tRNA，其他的酶因子以及能源。在体外系统中蛋白质合成的速率要比在体内慢一些。

mRNA 和小亚基结合，小亚基结合的 mRNA 的范围约 30 个碱基。但在任何的瞬间，肽键的合成只涉及到两个分子的 tRNA。因此多肽延伸反应只发生在核糖体所覆盖的 10 个密码子中的 2 个。

图 15－7 表明每个 tRNA 位于相应的位点，两个位点的特点不同。

(1)A 位点(或称进入位点，entrysite)可以进入氨基酰－tRNA。在 aa-tRNA 进入前，将加入到肽链中的下一个氨基酸的相应密码子出现在 A 位点。

(2)进入的新的氨基酸要被加到位于 P 位点(或称供位，donor site)的初生肽链上，P 位点是被肽基酰－tRNA 所占据，这个 tRNA 负载着新生的多肽链。

携着氨基酸的 tRNA 其末端位于大亚基，另一端的反密码子与结合在小亚基上的 mRNA 相互作用，因此 A 位点和 P 位点都必须延伸到大小两个亚基中（如图 15－7 所示）。核糖体合成肽链必须要在图 15－7 中的第一阶段的状态之下，即肽基酰－tRNA 处在 P 位点，氨基酰－tRNA 处于 A 位点。肽链的形成是通过肽基酰－tRNA 上的多肽的羧基转移到氨基酰－tRNA 上的氨基酸上这个反应来实现的。肽的转位是由核糖体大亚基的组分——肽转移酶催化的。

多肽链的转位产生了脱酰 tRNA(decrylated tDNA)，它仍位于 P 位点，但已不带有任何的氨基酸。而在 A 位点产生了新的肽酰－tRNA。这种情况表现在图中的第二步(图 15－7)，这个在 A 位点的肽酰－tRNA 要比原来在 P 位点的肽酰－tRNA 长一个氨基酸。

然后核糖体延着 mRNA 移动一个密码子。这一步称为移位(translocation)。这一移动使脱酰－tRNA 离开 P 位点，而肽酰－tRNA 进入 P 位点(见图 15－7，第三步)。在细菌中下一个新的密码子移入 P 位点后又进入了 E 位点，这个位点是在 tRNA 离开 P 位

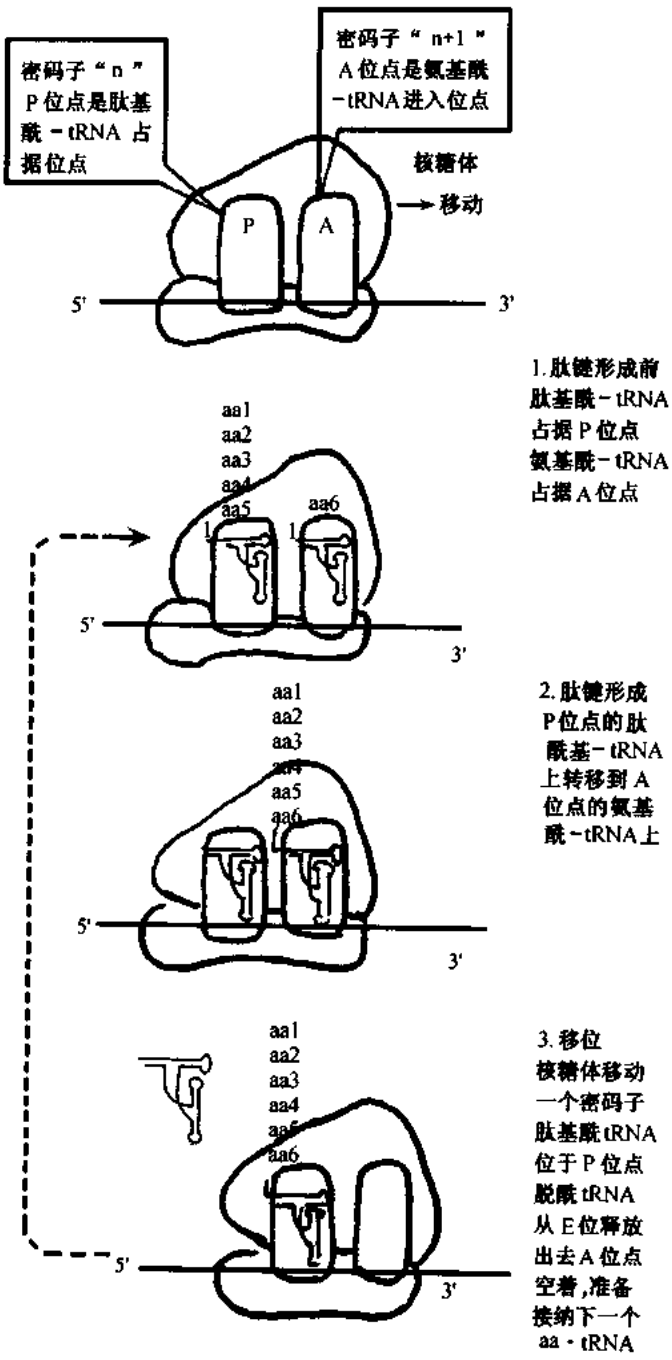


图 15 - 7　核糖体结合负载 tRNA 有二个步骤。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8 - 6)

点到释放出核糖体的途中被 tRNA 短暂地占据。此时 A 位点对氨基酰 - tRNA 的亲和性大大减低 (图 15 - 8、9)，阻止了另一个新的氨基酸的进入，使核糖体为接纳新的氨基酰 - tRNA 做准备。这样井然有序地循环往复。

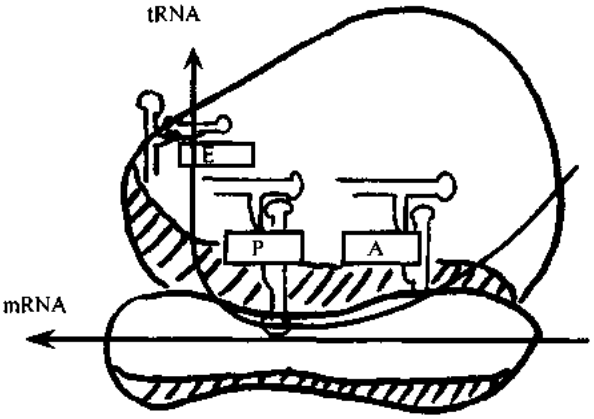


图 15－8　tRNA 和 mRNA 以相同的方向穿过核糖体。(仿 B.Lewin：Genes Ⅵ，1997，Fig.8－7)

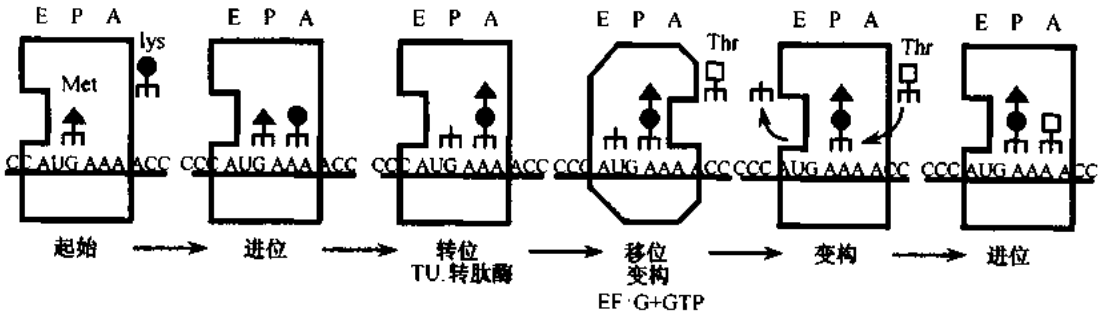


图 15－9　肽链延伸的三位点变构模型。

第二节　肽链的合成

一、细菌中翻译的起始

细菌核糖体以 70S 形式存在时承担着多肽链延伸的任务。翻译终止时它们仍以这种形式从 mRNA 上释放出来，然后进入到游离的核糖体库中。在细菌生长时，核糖体的主要任务是合成蛋白质。游离的核糖体库约保持在核糖体的 20%。

核糖体在游离库中可解离成分开的亚基，这样游离的 70S 核糖体与分开的 30S 和 50S 亚基以一种动态平衡的形式存在。完整的核糖体并不具有起始蛋白质合成的功能。起始是由分开的亚基来完成的，在起始反应中它们再重新聚集成完整的核糖体(图 15－10)。

在蛋白质合成起始时，mRNA 上有一个很短的保守顺序，称为**核糖体结合位点**(ribosome-binding site)起重要作用，此位点位于编码区前面也称为 **S－D 顺序**(Shine-Dalgarno sequence)。反应分为二步：①当小亚基在核糖体起始位点结合成起始复合物时，产生 mRNA 的识别；②然后大亚基结合到复合物上产生完整的核糖体。

这样小亚基和大亚基在 mRNA 上的核糖体结合位点聚集成完整的核糖体。虽然 30S 亚基涉及到蛋白质合成的起始，但它本身并没有结合 mRNA 和 tRNA 的这种反应能力。它需要称为起始因子(initiation factors，IF)的附加蛋白。这个因子最初是通过用

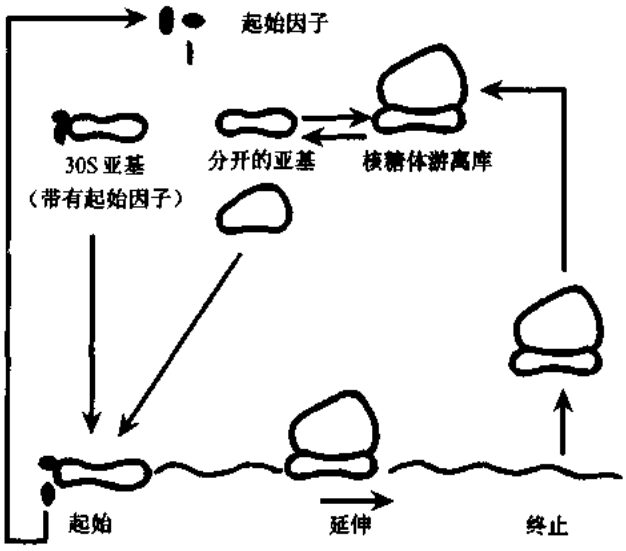


图 15－10　细菌核糖体在蛋白质合成中的循环。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8－9)

氯化铵洗涤 30S 亚基发现的，当通过这样的处理 30S 亚基就不能发动起始新链的合成。这是由于起始因子和 30S 亚基是以一种松散的结合，经氯化铵洗涤后它便从小亚基上释放出来。对核糖体蛋白不再有任何作用。起始因子仅结合在游离的 30S 亚基上，30S 的亚基来源于一般不参与蛋白质合成的核糖体池。每种因子的拷贝数要比游离核糖体数少得多。

表 15－2　*E. coli* 蛋白质合成起始所需的三种起始因子

因子	质量(kDa)	因子/核糖体	功能
IF3	23	25%	亚基解离与 mRNA 的结合
IF2	97.3	?	起始 tRNA 的结合与 GTP 水解
IF1	9	15%	循环因子?

B. Lewin: Genes Ⅳ. 1990, Table 7.2

当 50S 亚基与 30SmRNA 复合物聚集产生 70S 核糖体时，释放出起始因子。起始因子的丢失使起始从第一阶段转入第二阶段。这一行为表明起始因子不同于核糖体的结构蛋白。起始因子只和起始复合物的形成有关。在延伸阶段它们并不起作用。

细菌有 3 种起始因子 IF－1, IF－2 和 IF－3(表 15－2), mRNA 和 tRNA 进入起始复合体时是需要这些因子的。①IF－3 是 30S 亚基与 mRNA 起始位点的特异结合所必须的。②IF－2 是特异地和起始 tRNA 结合并把它带到起始复合体中。③IF－1 仅作为完整的起始复合物的一部分，可能和起始复合物的稳定性有关，而不是识别特异成分。图 15－11 表明了 IF－3 在起始中的作用。IF3 因子具有双重功能。

IF3 的第一个功能是控制了核糖游离态和结合态之间的平衡。IF－3 和游离的 30S 亚基相结合，但不和 70S 核糖体结合。当 IF－3 与 30S 亚基相结合时，阻止此亚基和 50S 亚基的结合，其作用实际上是抗缔合因子 (antiassociation factors)。在 IF－3 和 30S 亚基之间的反应是按一定的化学剂量的，每个亚基结合一个分子的 IF－3，因此 IF－3 的相对含量较少，这可能是由游离的 30S 亚基的量所决定的。

IF－3 的第二个功能是控制 30S 亚基与 mRNA 结合的能力，小亚基失去此因子就

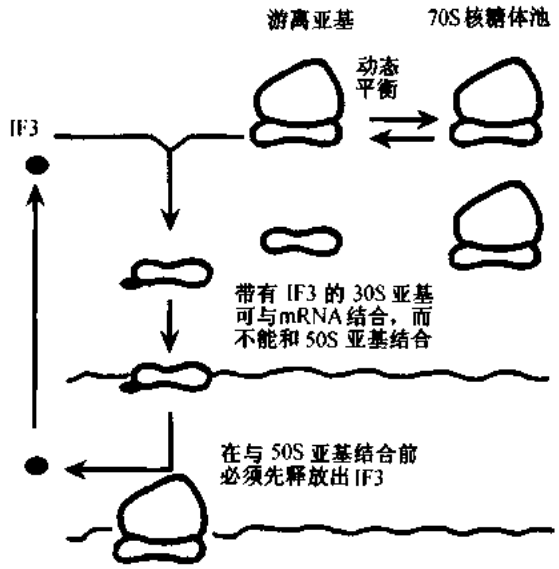


图 15－11　合成起始需要带有 IF3 的 30S 亚基。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8－10)

不能和 mRNA 形成起始复合体。但 mRNA 上特异位点的选择是核糖体亚基的功能。因此合成起始复合物需要 IF－3，但 IF－3 并不涉及特异位点的选择。

在 30S 亚基和 50S 亚基结合之前，IF－3 被释放出来。30S 亚基不能在结合 IF－3 的同时又与 50S 亚基结合。但这两种选择都是可以的。

(1)带有 IF－3 的小亚基因以游离的亚基形式存在，能和 mRNA 结合。

(2)当缺乏 IF－3 时，它可与 50S 亚基结合形成 70S 核糖体，承担翻译(若 30S 亚基已和 mRNA 结合)或者仍作为游离的核糖体(若 30S 亚基未和 mRNA 结合)。

二、起始 tRNA 的特点

翻译起始时基因的第一个密码子是怎样被识别的呢？这个反应分为二个阶段。

(1) 细菌 mRNA 上的核糖体识别位点的顺序是由带有 IF－3 的 30S 亚基识别的。这个由核糖体直接识别的位点称为 Shine-Dalgarno 顺序，它在 AUG 上游，由 10 个碱基组成(5′－AAACAGGAGG－3′)和小亚基上 16SrRNA3′端的六核苷酸(3′－UCCUCC－5′)正好互补，相互结合，使下游的 AUG 起始密码子定位在 P 位上。

(2) 核糖体结合位点也含有一个信号，产生开始的读框，这是特异的**起始密码子**(initiation codon)，通常它是三联体 AUG，但在细菌中 GUG 或 UUG 也被同样作为起始密码子。

AUG 是相对于 Met 的密码子，而两种类型的 tRNA 都能携带 Met，但一个是用于起始，另一个是用于延伸。

在细菌和真核细胞器中起始 tRNA 携带着一个甲酰化的甲硫氨酸残基，形成了**N－甲酰甲硫氨酸－tRNA**(N-formyl-methionyl-tRNA)分子。此 tRNA 称为 $tRNA^{fMet}$。氨基酰－tRNA 的名字通常缩写成 $fMet\text{-}tRNA_f$。

在二步反应中起始 tRNA 得到了已修饰的氨基酸。首先是负载上氨基酸产生了一个 $Met\text{-}tRNA_f$。然后通过甲酰化反应封闭了自由的 NH_2 基(图 15－12)。虽然被阻断的氨基可能抑制起始子参与链的延伸，但并不影响它起始合成蛋白的能力。

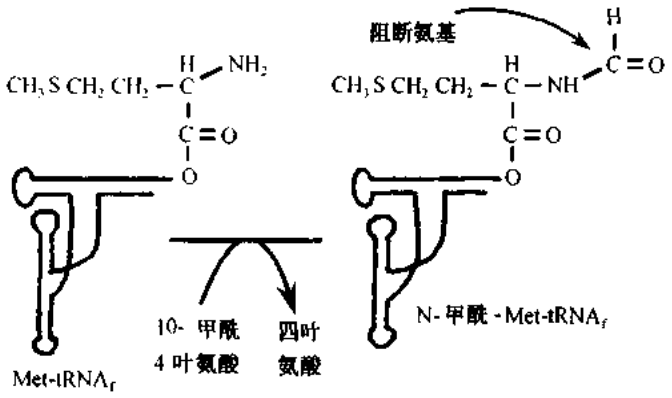


图 15 - 12　起始子 N - 甲酰 - Met - tRNA 是由甲硫氨酰 tRNA 产生的。
(仿 B.Lewin: Genes Ⅵ, 1997, Fig.8 - 11)

这种 tRNA 仅用于起始，它识别密码子 AUC 或 GUG(偶尔也识别 UUG)。这些密码子的识别并非是同等的。当 AUG 被 GUG 取代时起始的效率会下降一半，若用 UUG 时起始程度又要下降一半。

负责识别延伸时的 AUG 密码子的是 $tRNA_m^{Met}$，这个 tRNA 的甲硫氨酸没有被甲酰化，它仅负责非起始的 AUG 密码子。

起始和延伸的 Met - tRNAs 有两点区别：①两种 tRNA 本身有差别；②氨基酸的状态不同。

AUG 和 GUG 密码子的意思取决于“上下文”关系，当 AUG 用于起始时，它是作为甲酰 - 甲硫氨酸的密码子。GUG 的意思主要取决于其位置。当它位于第一个密码子时它被读成为甲酰 - 甲硫氨酸，参与起始反应。当它位于基因中时，它被 Val-tRNA 所识别(表 14 - 2)。

AUG 和 GUG 密码子的“上下文”如何解释呢？图 15 - 13 表明核糖体的作用。在一个起始复合物中 mRNA 的起始密码子正好位于小亚基的 P 位点上，只有 fMet-tRNA 才能成为起始复合物的一部分，它就称为**起始子**（initiator），其惟一的特点就是可以

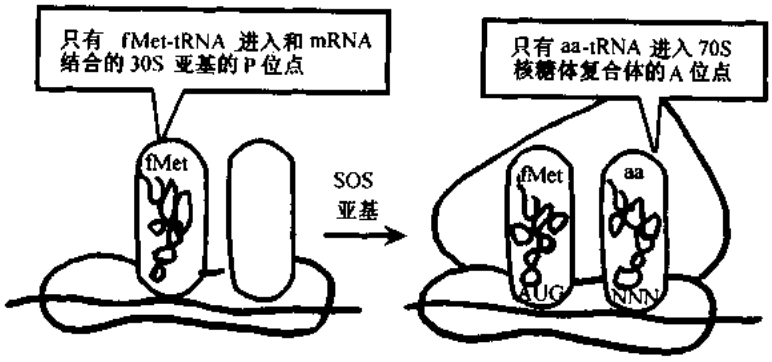


图 15 - 13　只有 fMet-tRNA 被 30S 亚基用于起始，只有另一种氨基酰 - tRNA 被 70S 核糖体用于延伸。(仿 B.Lewin: Genes Ⅵ, 1997, Fig8 - 12)

直接进入 P 位点，来识别其相应的密码子。当大亚基加入到复合物中，这个起始子 $fMet\text{-}tRNA_f$ 位于新的完整的 P 位点，而与第二个密码子互补的氨基酰 - tRNA 可进入A 位点。第一个肽链在起始子 tRNA 和第二个氨基酰 - tRNA 之间形成，此起始子

tRNA 的行为是作为肽酰－tRNA 的类似物（此类似物的意思是指它的肽链上只含有一个氨基酸。）

当 AUG(或 GUG)密码子位于核糖体 P 结合位点中时，起始有效进行，因为只有起始 tRNA 才能进入 30 亚基和 mRNA 结合时所产生的 P 位点。而当 mRNA 的密码子进入核糖体时，主要是翻译内部的阅读框，因为只有氨基酰－tRNA 才能进入 A 位点。

起始的 fMet-tRNA 与 Met-tRNA$_m$ 以及其他用于延伸的 tRNA 有什么不同的特点呢？甲硫氨酸的修饰将阻止 fMet-tRNA 用于延伸，但这种修饰不是作为起始 tRNA 所严格要求的，因为 Met-tRNA$_f$ 也具有起始子的功能。tRNA 顺序的一些结构特点是两种 tRNA 的重要差别(图 15－14)。其中有的特点是延伸所必要的。

(1) 在受体臂上最末端的一对碱基除 tRNA$_f^{Met}$ 以外都是配对的，在 tRNA$_f^{Met}$ 这个位点若发生突变产生碱基对的话，那么它将不用于起始而用于延伸。因为 C 和 A 不配对时，C 才可以和延伸因子 IF－2 相结合。

(2) tRNA$_f^{Met}$ 在反密码子环的茎上有 3 对 G－C(图 15－14)，若发生突变将会阻止其进入 P 位点。

(3)tRNA$_f^{Met}$ 的反密码子环上的是 A，tRNA$_f^{Met}$ 是烷基化腺嘌呤。前者的 TΨC 环上 A

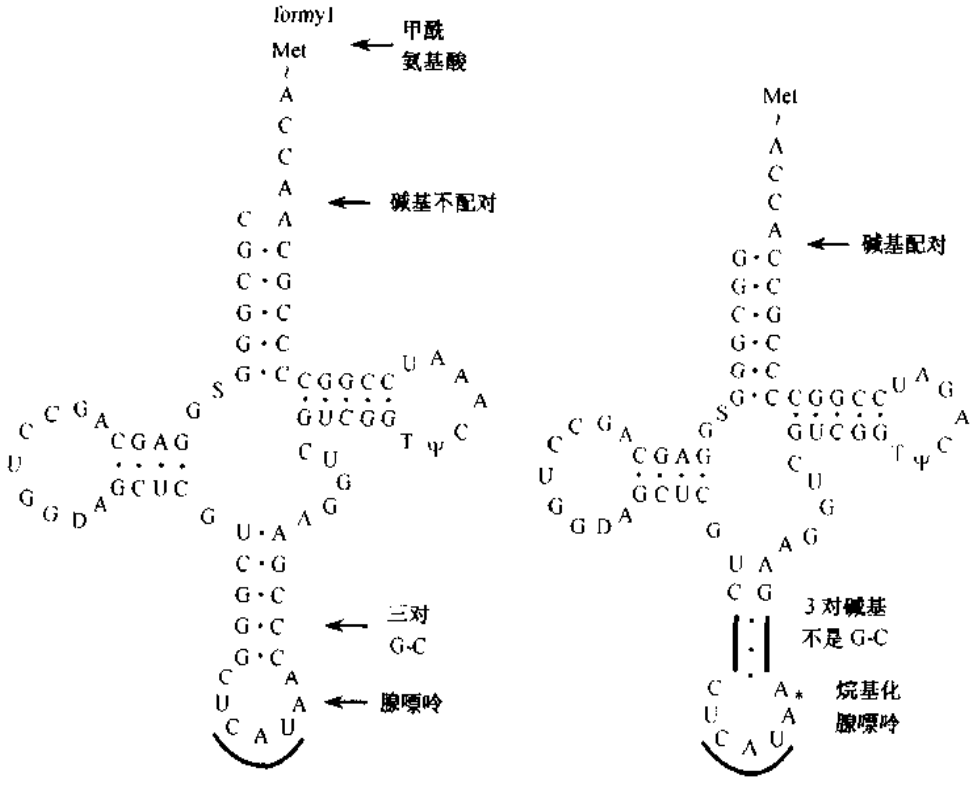


图 15－14　fMet-tRNA$_f$ 和 Met-tRNA$_m$ 的区别。(参考 B. Lewin: Genes V，1994)

的相应位置上 tRNA$_f^{Met}$ 是 G，它们的作用尚不清楚。

氨基酰－tRNA 进入核糖体的能力是受 IF－2 因子控制的。fMet-tRNA 起始因子是由 IF－2 结合到 30S mRNA 复合体上的(图 15－15)首先 IF－2 和 30S mRNA 复合体结合，同时 GTP 和 30S mRNA 复合体结合，然后 fMet－tRNA 通过 IF－2 结合到 30S mRNA 复合体上，形成起始复合体，然后 IF－2－fMet－tRNA 二元复合体再和 30SmRNA 复合体结合，起始 tRNA 位于 P 位点。

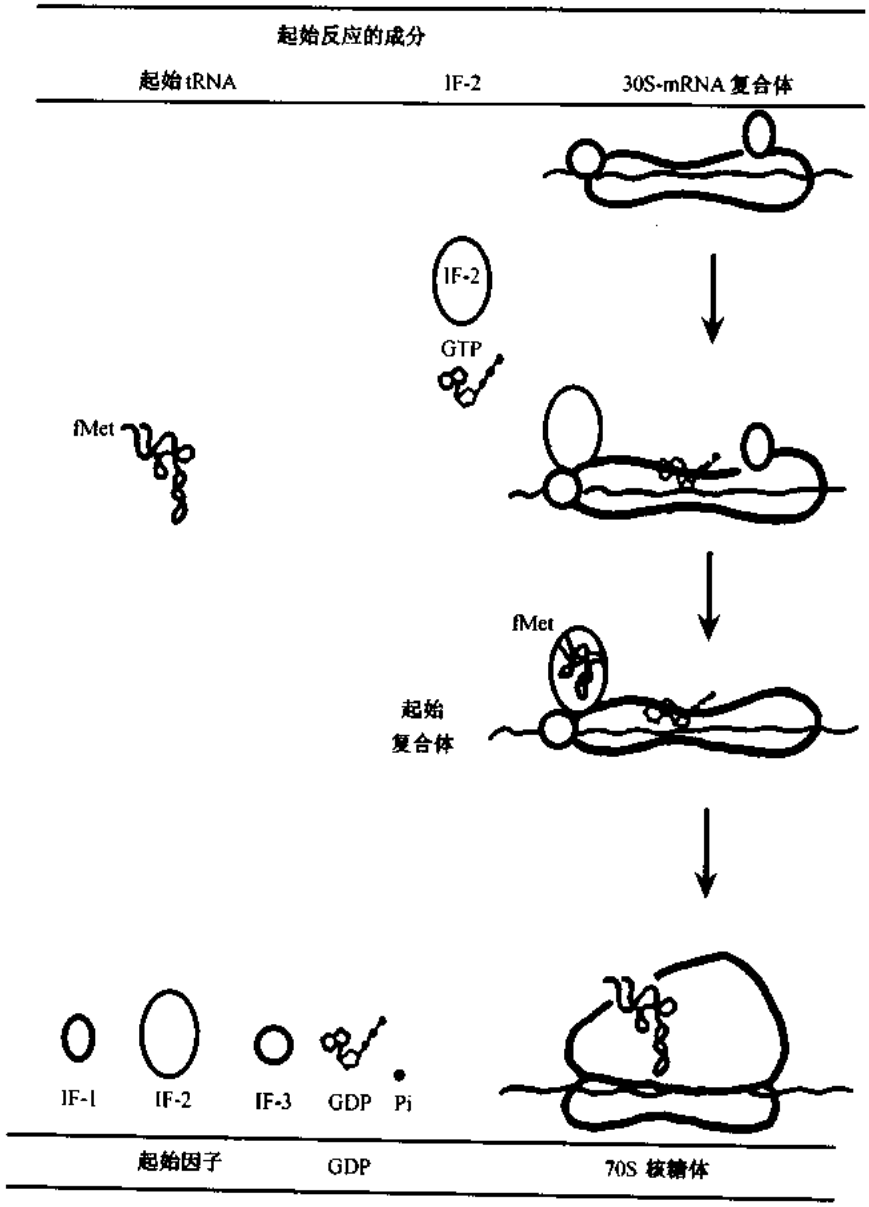


图 15－15　原核生物翻译的起始。IF－2 对于 fMet-tRNA 和 30S-mRNA 复合体的结合是必要的，50S 亚基结合后，所有的 IF 因子却被释放出来，GTP 被水解。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8－14）

只有起始 tRNA 才能和 IF－2 结合。别的任何氨基酰 tRNA 都不能参加起始反应。

在这一阶段，IF－2 保留在 30S 亚基上，可以进一步地发挥作用。这个因子具有核**糖体依赖性 GTP 酶活性**（ribosome-dependent GTPase activity）：它负责水解存在于核糖体上的 GTP。释放出高能键中储存的能量。

当 50S 亚基和 30S 亚基结合形成完整的核糖体时，GTP 已被水解。IF－2 本身并不是 GTPase，但可以激活带有这种功能的核糖体蛋白。GTP 的水解可能涉及到核糖体构象的改变，这样结合后的亚基变成了有活性的 70S 核糖体。

由于同一个氨基酸被用于所有蛋白链的起始，有时需要切除掉这个不需要的残基（图 15－16），这个反应发生很快，当新生肽链达到 15～16 氨基酸长度时就会发生。

在细菌和线粒体中，甲酰基是通过一种特殊的**去甲酰化酶**（deformylase）产生一个正常的 NH2 末端，若甲硫氨酸是蛋白质 N－端的氨基酸，那么只需要这一步反应。约

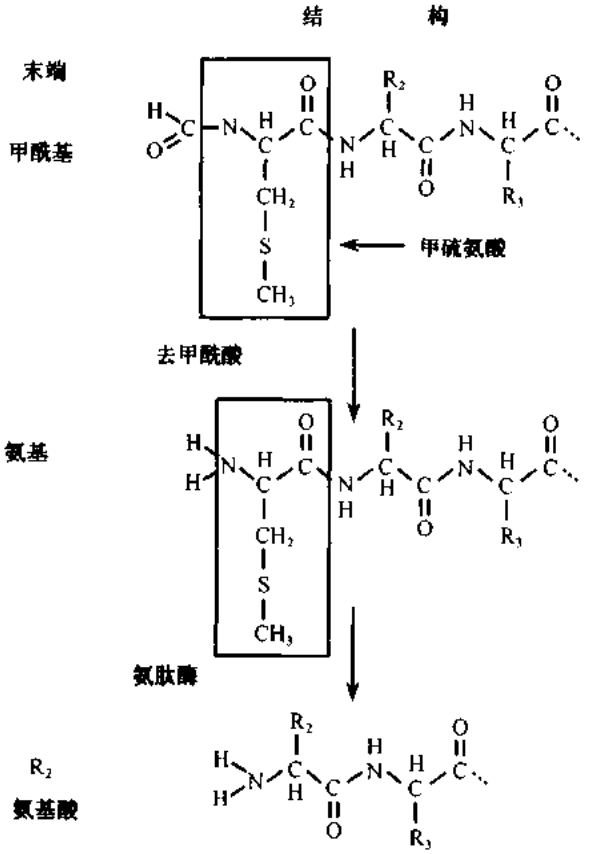


图 15－16 细菌新合成的蛋白开始是有 fMet，但甲酰化 Met 在蛋白合成后可能会被除去。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8－15）

有半数的蛋白质其末端的甲硫氨酸再被**氨肽酶**（aminopeptiase）切除掉，产生一个新的 R2 末端（图 15－16）。当成熟肽链 N－端的第一个氨基酸不是甲硫氨酸时，这两步反应都需要。

三、真核生物蛋白质合成的起始

在真核生物细胞质中涉及到蛋白质合成起始的组分和原核生物是相似的，密码子 AUG 总是用于起始子；用 GUG 作为起始密码很少发现。起始子 tRNA 是很特殊的，可是它携带的甲硫氨酸并不被甲酰基化，称为 $tRNA_i^{Met}$ 这样一来起始和延伸的 Met-tRNAs 之间的区别仅在 tRNA 部分，Met－$tRNA_i$ 用于起始，而 Met-$tRNA_m$ 则用于延伸。

细菌的 30S 和真核的 40S 亚基在起始蛋白质合成时与 mRNA 结合的位点是不同的。在细菌中直接在起始密码子 AUG 附近形成起始复合体。而在真核中小亚基首先识别 *m*RNA 的 5′端，然后再移到起始位点，和大亚基结合。这个细节在后面将会介绍。

除了这些差别以外真核生物蛋白合成的起始的过程和 *E. coli* 的相似，但有更多的起始因子（表 15－3）。这些因子的名称和原核中的相似，但在前面加上“e”，表示是真核生物的起始因子。

在真核起始过程中，由 Met-$tRNA_i$，eIF-2 和 GTP 形成了**三元复合体**（ternary complex）。此复合物的形成分为二步：首先 GTP 和 eIF-2 结合，以增加此因子与 Met-$tRNA_i$。

表 15－3　真核生物蛋白质合成中涉及的各种辅助因子

	因子	分子量(kDa)	结构	功能
起始	eIF－3	550	多聚体	40S 三元复合体和 mRNA 结合，和 Cap 有强的亲和力
	eIF－4F(CBP－Ⅱ)	220	多聚体	5′帽结合蛋白，具有解链酶活性
	eIF－4E(CBP－Ⅰ)	24	单体	结合 mRNA5′端，解链
	eIF－1	15	单体	帮助 mRNA 的结合，形成 40S 起始复合物
	eIF－4B	80	单体	结合 mRNA，解链酶
	eIF－4A	44.4	单体	结合 mRNA 和 ATP，水解 ATP，解链
	eIF－6	23	单体	阻止 40S 和 60S 亚基结合
	eIF－5	150	单体	介导 eIF－2 和 eIF－3 从起始复合体中释放出来
	eIF－4C		单体	40S 和 60S 亚基结合
	eIF－2α	35	三体	结合 GTP，由磷酸化控制
	eIF－2β	38		可能是循环因子
	eIF－2γ	55		结合 Met-tRNA$_f$
	eIF－1A	17.5	单体	核糖体解聚，结合 60S 亚基
	eIF－3A	25	单体	核糖体解聚，结合 60S 亚基
延伸	eIF－5A	16.7	单体	促进第一个肽链的形成
	eEF－1α	51	单体	结合 aa-tRNA，GTPase
	eEF－1β	23	单体	与 eEF－1α 的 GTP:GDP 交换
	eEF－1γ	49	单体	GTP:GDP 交换
	eEF－2	100	单体	依赖于 GTP 水解的移位酶，相当于原核的 EF-G
终止	eRF	54	单体	识别 3 种终止密码，促进脱酰－tRNAd 的释放

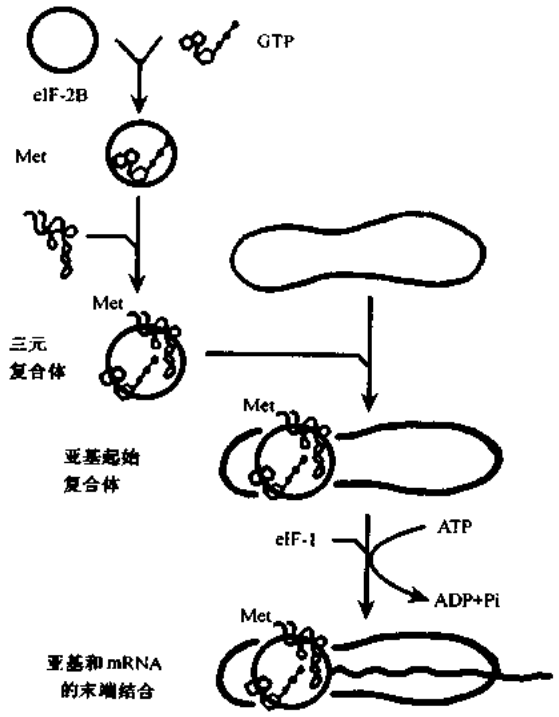


图 15－17　真核生物蛋白质合成起始。(参考 B.Lewin: Genes Ⅴ, 1994, Fig.7－20; Genes Ⅵ, 1997, Fig.8－18)

结合的效率。然后三元复合体再与 40S 亚基结合(图 15－17)。此反应发生在和 mRNA 结合之前。实际上 Met-tRNA$_i$ 起始子必须按一定的顺序与 mRNA 结合，因此 eIF-2 对应于原核的 IF-2，但它和小亚基的相互作用发生在起始的早期阶段。

mRNA 的识别需要几种附加因子；它们主要的功能是除去 mRNA 中的二级结构。起始因子的作用还涉及到亚基起始复合体和 mRNA 的结合。

eIF－4 因子是一种多体，包括 CBP(cap binding protein) 和 eIF-4A，CBP 识别 mRNA5′端的“帽子”而 eIF-4A 可使 mRNA 前端 15 个碱基中任何二级结构解链，解链的能量依赖于水解 ATP 来供给。eIF-4A 和 eIF-4B 的共同作用进一步延着 mRNA 解链。

40S 三元复合体和 mRNA 的结合依赖于 eIF-3，它并不像细菌中的 IF-3，主要功能是和 mRNA 的结合有关。eIF-3 是一个很大的因子，带有 8－10 个亚基。IF-3 结合在小亚基上。

当小亚基和 mRNA 结合时通常向第一个真核密码子 AUG 移动，关于这个过程知道的很少。推测它需要从 ATP 水解中获能量。当小亚基到达起始位点时便停下来并且

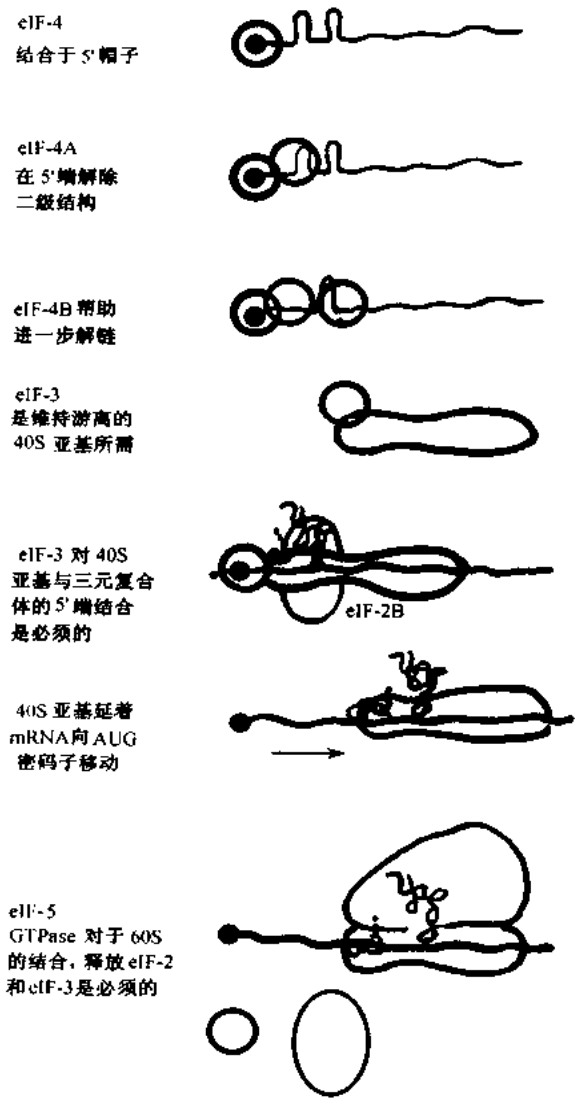


图 15－18　各种真核生物起始因子对解开 mRNA 二级结构，小亚基与起始复合体的结合及支持大亚基的结合都是必须的。(仿 B.Lewin：Genes Ⅵ，1997，Fig.8－19)

能和大亚基相结合。

eIF－3 和 eIF－6 是在维持亚基处于解离状态中所必要的。IF－6 结合在大亚基上，它可能作用大亚基，当大亚基和小亚基结合成起始复合体时它被释放出来。

60S 亚基和起始复合体的连结是在 eIF－2 和 eIF－3 释放之后，eIF－5 的功能就是介导 eIF－2 和 eIF－3 从起始复合体中释放出来。eIF－5 是一个 GTPase，它水解 GTP，对于大、小亚基的结合和 eIF－2、eIF－3 的释放都是必要的。当 80S 完整的核糖体形成时，各种因子都被释放出来(图 15－18，19)。

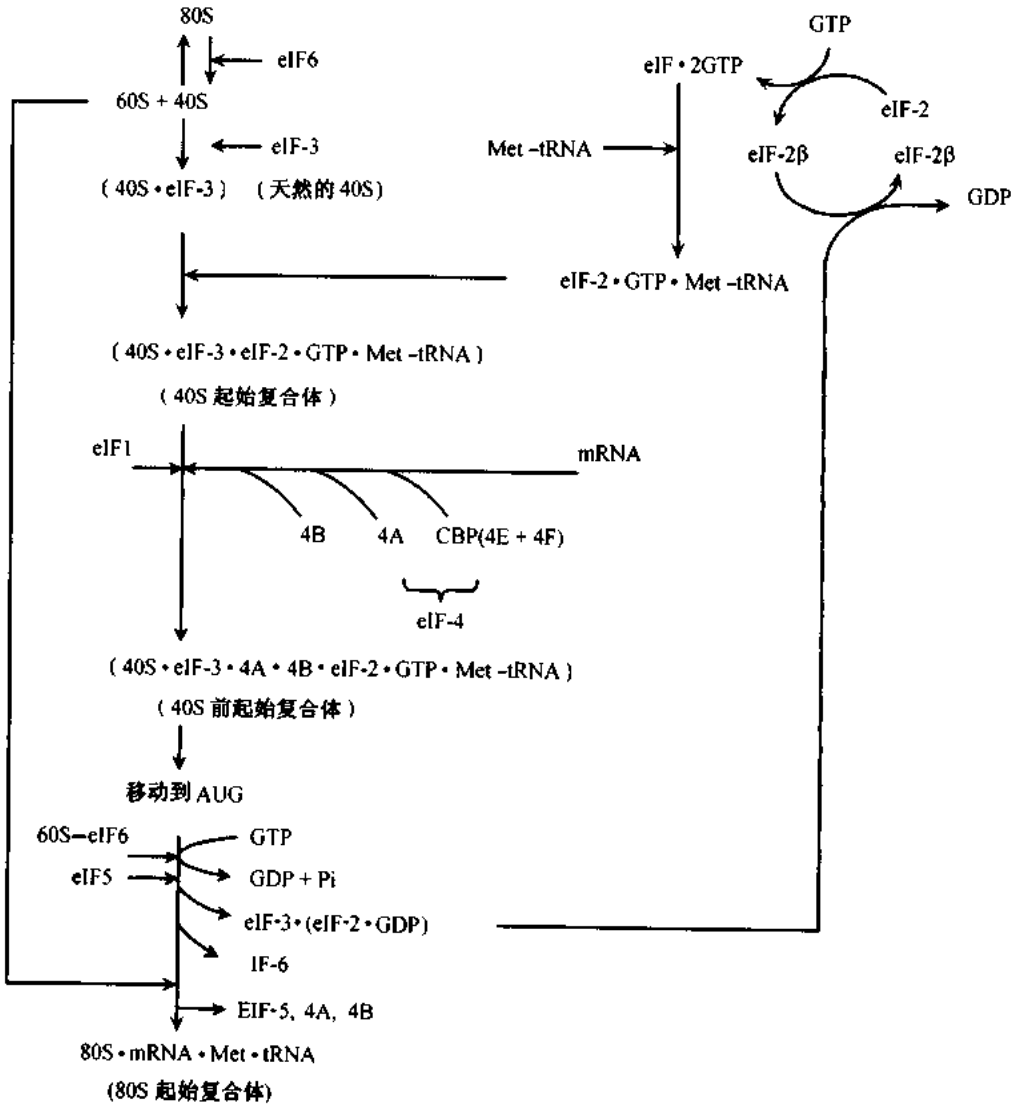


图 15－19　真核生物蛋白质合成中 80S 起始复合物的形成。

四、起始时 mRNA 和 rRNA 的碱基配对

1. 原核 30S 小亚基和 S－D 顺序的结合

mRNA 上蛋白质合成起始位点可在阻断延伸的条件下通过核糖体与 mRNA 的结合来鉴别。此时核糖体保留在起始位点上，加入核酸酶，使核糖体附着区的以外区域都被降解，而结合的区域因受保护而被留下来。这一片断通过分析可了解其结构和特点。

在细菌中被核糖体保护的真核区长 35～40 碱基长。在不同的细菌中这段顺序同源性很弱，它们只有 2 个共同的特点。①这一区域中总含有起始密码子 AUG(是 GUG 或 UUG 的情况较少)；②有一个顺序和 16S rRNA 近 3′端的顺序互补。

以反方向写的话 16S rRNA 近 3′端是以下的 6nt 组成的保守顺序：

3′……UCCUCC……5′

已知 *E.coli* 的各种 mRNA 起始位点至少有 3 个碱基与其互补，通常是 4～5 个碱基。这段顺序是

5′……AGGAGG……3′

此多聚嘌呤的顺序称为 Shine－Dalgaruo(S－D)顺序，它位于真核密码子前约 7 个碱基。

在 mRNA 和核糖体结合中 S－D 顺序和 rRNA 配对起什么作用呢？经点突变试验表明它们的配对在起始中有着重要的作用。S－D 顺序的点突变能阻止被翻译。如在 rRNA 的互补顺序上引入突变将对细胞有害，它会改变蛋白质合成的模式。为了进一步搞清碱基配对反应，在 mRNA 的 S－D 顺序和 rRNA 进行了补偿性的突变，结果所产生的 mRNA 翻译起始的缺陷特异地被相应突变的核糖体所校正。

细菌的 16S rRNA 的 3′端顺序是高度保守的。它可以自我互补，形成发夹结构。发夹的有一部分可以和 mRNA 的 S－D 顺序互补。

这段顺序不可能同时既形成发夹结构又和 mRNA 结合，在起始反应时可能末端的发夹打开而与 mRNA 配对(图 15－20)。起始以后，mRNA-rRNA 的双链又因发夹结构重新形成而断开。这一机制可能服从在特殊位点形成一个稳定起始复合物，然后此复合物又要沿着 mRNA 移开的需要。

rRNA 3′端顺序在原核与真核都存在，在双甲基腺嘌呤附近和 3′端之间约 20 个碱基，但有两点重要的区别，在细菌中此处有两个 U，而在高等真核中都是两个 A(在低等真核生物中是属中间型的 AU)，在所有的真核生物中都缺少了 CCUCC 和 S－D 顺序互补的 5 个碱基。

有很多的间接证据表明：在真核的 rRNA 中可能有一个 S－D 顺序相应的其他顺序叫做 Kozak 顺序，序列为 ACCAUGG，但现在还未证实真核 mRNA 和 18S rRNA 之间存在碱基配对。若实际上没有这个机制的话，这就构成了两者之间起始机制的重要差别。

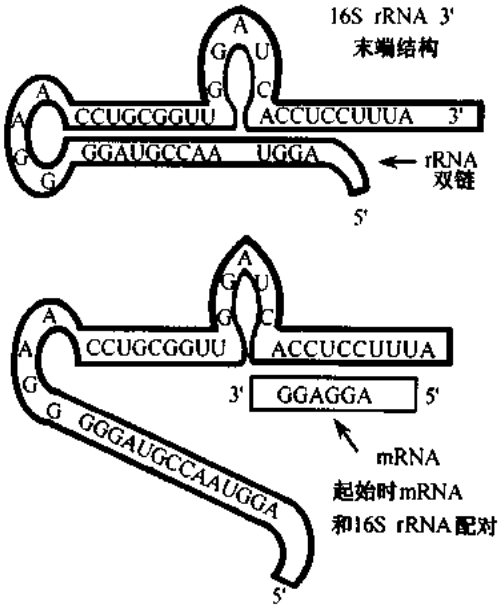


图 15－20　16S rRNA 3′端的发夹结构及和 mRNA 的互补。(仿 B. Lewin: Genes V, 1994, Fig. 10－12)

2. 真核合成起始时小亚基和起始位点的结合

各种真核 mRNA 都比编码其蛋白所需的 mRNA 长度要长，这是由于编码区仅占据所有信息的一部分。在真核细胞质中 mRNA 平均长度为 1～2kb，在 5′端有甲基化的帽子，3′端有 100～200nt 的 poly(A)尾巴，5′端不翻译的前导顺序(leader)相对较短，一般小于 100nt 长，编码区的长度是以相应蛋白质的大小而定。5′端不翻译的拖尾顺序(trailer)一般较长，有时可达 1000 碱基。根据它们的位置的特点，前导顺序在起始中的作用毋容忽视，至于拖尾顺序我们还不知道它对起始有什么作用。

真核细胞质中的核糖体并不在编码区开始处直接和起始位点相结合，而首先识别 5′端甲基化的帽子结构。若除去帽子结构这个信息，在体外便不能有效地进行蛋白合成。40S 亚基和 mRNA 的结合需要各种因子，包括那些识别帽结构的蛋白质。

我们已论述了起始的过程，表明二级结构可能妨碍了与核糖体结合位点的结合效率，减少了对起始的识别能力。帽结合蛋白(CBP)的功能之一就是使前导顺序解链，来帮助核糖体结合位点以单链的状态存在，以保证有效识别。在图 15－18 中表明真核起始因子中 eIF4A 和 eIF4B 都具有使 mRNA 的前导顺序二级结构解链的作用。

5′端的修饰几乎发生在所有的真核细胞和病毒 mRNA 上且对于翻译是十分必要的。唯一的例外是少数病毒的 mRNA[如脊髓灰质炎病毒 (poliovirus)，它是未加帽子的]。这些例外的病毒 mRNA 在体内翻译不需要帽结构，它们具有不需要帽子而进行翻译的一些位点，叫做内部核糖体进入位点(internal ribosome entry site, IERS)。此位点是一种可以形成二级结构的模体，可使蛋白质合成从内部起始。在果蝇触角足复合基因座(见 24 章)也存在此位点。

有些病毒利用了这些差异。如脊髓灰质炎病毒感染抑制了宿主 mRNA 的翻译，这是通过阻碍细胞 mRNAs 起始所必需的 CBP 来实现的，而脊髓灰质炎病毒本身的无帽结构的 mRNA 就不会受到影响。

有时 AUG 起始密码是位于 mRNA 5′端 40 个碱基处，因此帽子和 AUG 同处于核糖体的结合区中。但也有很多的 mRNA 的 Cap 和 AUG 相隔有一段距离，最长的可达 1000 碱基。帽子存在对于在起始密码子处形成稳定的复合物仍然是必要的。核糖体怎样能在两个相隔很远的位点上发挥作用呢？图 15－21 描绘了扫描(scaning)模型，认为 40S 亚基开始识别 5′端的帽子，然后再延着 mRNA 移动。从 5′端开始扫描是一个线性的过程。当40S 亚基扫描前导顺序时，他们使稳定性小于 30 千卡的发夹二级结构解链。但稳定性大的发夹就会妨碍或阻止核糖体的迁移。

当 40S 亚基迁移到 AUG 起始密码子时，迁移就停下来。通常（但不总是如此）第一个 AUG 三联体就是起始密码子。此 AUG 三联体本身并不足以使迁移停下来。很可能当具有正确的邻接顺序时它作为起始密码子提供有效识别。最适合的邻接顺序是 GCCDDAUGG，AUG 前面的第三个嘌呤（AorG）和后面的一个 G 是最重要的，它影响转录的效率达 10 倍。其他的碱基作用较小。当前导顺序较长时更多的 40S 亚基能识别第一个起始密码子前面的 5′端，延着前导顺序核糖体排成一排，直至起始位点（图 15－21）。

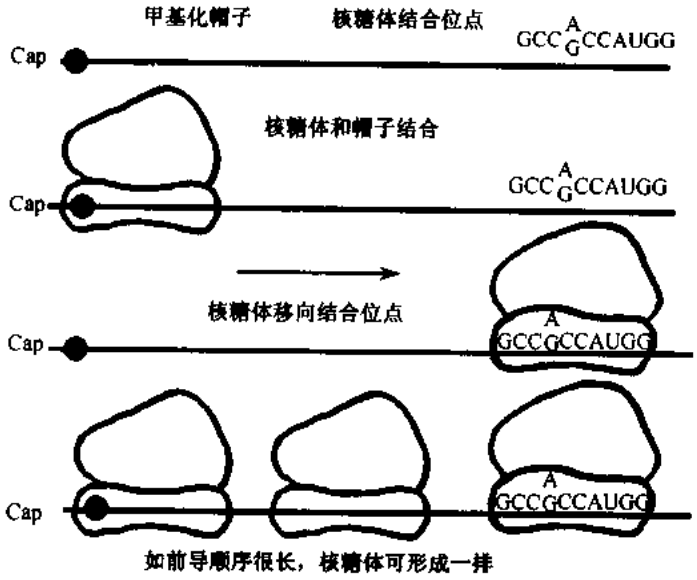


图 15－21　真核核糖体从 5′端向结合位点移动。(仿 B.Lewin：Genes Ⅵ，1997，Fig.8－17)

在起始位点上的结合是稳定的。当 40S 亚基和 60S 亚基连接起来形成完整的核糖体，其所处的位点可以通过保护的方法加以鉴别。一个 40S 亚基保护的区域达到 60 碱基，而形成完整的核糖体后其保护区域和原核相似约 30～40 个碱基。

核糖体必须从 mRNA 5′端开始运作，这是和真核 mRNA 为单顺反子的特点是一致的。某些例外的情况是一些病毒的 mRNAs 中，它含有一个以上的编码区，在这些情况下，靠 mRNA 编码区最近的 5′端先被翻译。其他顺序只有在此 mRNA 被剪切后，在中部起始密码子处产生一个新的 5′端才能被阅读。此现象表明，中部的起始位点不能被直接识别。

第三节　肽链合成的延伸和终止

一、肽链合成的延伸

肽链的延伸可分为三个阶段：① **进位反应**：主要是密码子－反密码子的识别；② **转位反应**：涉及肽链的形成；③ **移位反应**：tRNA 和 mRNA 相对核糖体的移动。在细菌中涉及到三位点模型。

1. 进位反应

一旦在起始密码子上形成了完整的核糖体。就开始进位反应阶段，氨基酰－tRNA 进入核糖体的 A 位点，而肽酰－tRNA 占据 P 位点，这样不断往返循环。除了起始子以外，任何的氨基酰－tRNA 只能进入 A 位点。它的进入是由**延伸因子**(elongation factor)(在细菌中是 EF－Tu)(表 15－4)的参与才得以完成的。

正像起始中的 IF-2 一样，EF-Tu 仅在氨基酰－tRNA 进入核糖体时，它才和核糖体

表 15－4　原核生物蛋白质合成的延伸因子

因子	基因	分子量(Da)	分子数/细胞	功能	抑制剂
EF Tu	*tufA*，*tufB*	43 225	70 000	与氨基酰 tRNA 及 GTP 结合	黄色霉素
EF Ts	*tsr*	74 000	10 000	结合 EF-Tu，取代 GDP	
EF G		77 444	70 000	结合核糖体和 GTP	梭链孢酸

B. Lewin: Genes Ⅳ, 1990, Table 7.5 和 7.7

结合，一旦氨基酰－tRNA 到位了，那么 EF-Tu 被核糖体释放出来，它再与另一个氨基酰－tRNA 结合。它显示出与核糖体结合和解离的循环。

图 15－22 表示氨基酰 tRNA 进入 A 位点的途径。EF-Tu 带有一个鸟苷，这个因子是由鸟苷的状态控制蛋白质活性的又一个例子。①当 GTP 存在时，EF-Tu 呈活性状态。②当 GTP 水解成 GDP 时，EF-Tu 便失活。③当 GDP 被 GTP 取代后，它又恢复了活性。

EF-Tu·GTP 的二元复合体与氨基酰－tRNA 结合成氨基酰－tRNA·EF-Tu·GTP 三元复合体。此三元复合体只有当 P 位点已被肽酰－tRNA 占据时才能和 A 位点结合。这是一步使氨基酰－tRNA 和肽酰－tRNA 处于正确位置进而形成肽链的关键性反应。

氨基酰－tRNA 位于 A 位点后 GTP 被裂解，同时释放出 EF-Tu·GDP。此时的 EF－Tu 是处于失活状态，不能和氨基酰 tRNA 结合。

另一个因子 EF-Ts 会导致失活型的 EF-Tu·GDP 变成为有活性的 EF-Tu·GTP。首先 EF-Ts 从 EF-Tu 中转换出 GDP，形成 EF-Tu·EF-Ts 复合因子，然后 EF-Ts 再被 GTP 取代，重新形成 EF-Tu·GTP 释放出 EF-Ts 进入再循环。此有活性的二元复合体和氨基酰－tRNA 结合。

延伸因子 EF-Tu 在游离态和结合态之间进行循环。

每个细菌中约有 70 000 个分子的 EF-Tu，接近于氨基酰 tRNA 的数量，这就意味着大部分的氨基酰－tRNA 可能都以三元复合体形式存在。EF-Tu 由两个基因编码，提供

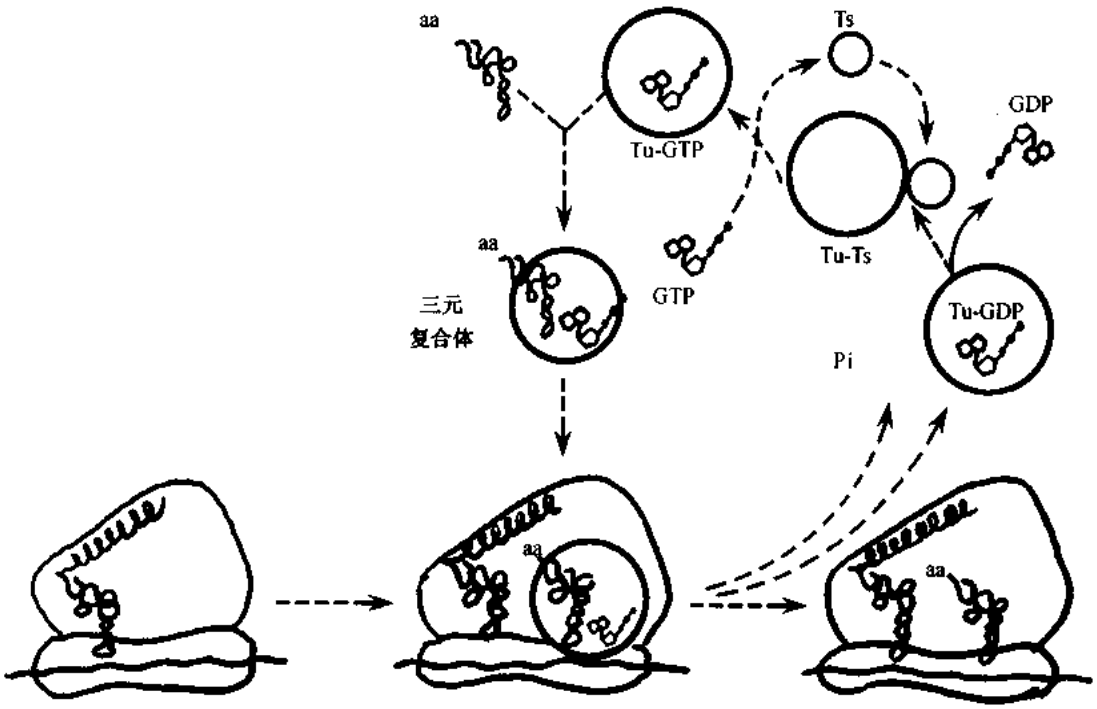


图 15－22　原核生物肽链延伸的进位反应。(参考 B.Lewin:Genes Ⅵ,1997,Fig.8－20)

了 5%的细菌总蛋白。大部分的 EF-Tu 都形成二元或三元复合体，因每个细胞中 EF-Ts 只有约 10 000 个分子。

在三元复合体中 GTP 的作用已通过用一个不能水解的类似物来取代它进行了研究。这个类似物是 GMP-PCP，它在氧的位置有一个甲硫氨酸桥，将 GTP 的 β 和 γ 磷酸连接起来。含有 GMP-PCP 的三元复合体也能和氨基酰－tRNA 结合，将其带到核糖体中，但肽链不能形成。因此 GTP 的存在看来对于氨基酰－tRNA 在 A 位点的结合是必要的。但水解 ATP 对此过程是不必要的。

黄色霉素(kirromycin)是一种抗生素，它可以抑制 EF-Tu 的功能。当 EF-Tu 被黄色霉素结合时，它仍能和氨基酰－tRNA 结合到 A 位点，但 EF-Tu·GDP 复合体不能从核糖体中释放出来。它们的持续存在阻止了肽酰－tRNA 和 aa-tRNA 之间肽链的形成。这样导致核糖体停留在 mRNA 上，蛋白质的合成也就停止了。

黄色霉素的作用表明抑制蛋白质合成的一个步骤就会打断下一个步骤，EF-Tu·GDP 的释放对于核糖体产生一个负责肽链形成的正确的构象是必要的。同理蛋白合成的其他阶段也是如此：在下一个反应发生之前，前一个反应必需全部完成。

在真核生物中，延伸因子 eIF-1 负责将 aa-tRNA 带到核糖体中，进一步裂解 GTP 的高能键，这种激活因子是由不同大小的多肽链集合而成。GTP 裂解后如何重复产生，细节尚不清楚。此因子的组成可能和 EF-Tu 及 EF-Ts 相似。从数量方面来说 eEF-1 在真核细胞中的分布和 EF-Tu 在原核中的分布是相似的，构成了细胞中主要的蛋白。

在真核肽链合成的起始中结合于 eEF-2 的 GDP 必需被 GTP 所取代，然后 eEF-2 才能被使用。这种鸟苷酸交换反应对于触动各种调节反应是很重要的，它将细胞中整个的翻译的速率和其他代谢途径连接起来。

2. 转位反应

当多肽链从 P 位点的肽酰 - tRNA 上转移到 A 位点的 aa-tRNA 上时，核糖体仍保持在原来的 mRNA 位置上。图 15 - 23 就表示这个反应。负责肽链合成的活性被称为**肽转移酶**(prptityl transferase)。

转位反应的特点通过抗生素嘌呤霉素（putomycin）抑制蛋白质合成的能力所显示。嘌呤霉素和附着在 tRNA 末端腺嘌呤的一个氨基酸很相似。图 15 - 24 表示嘌呤霉

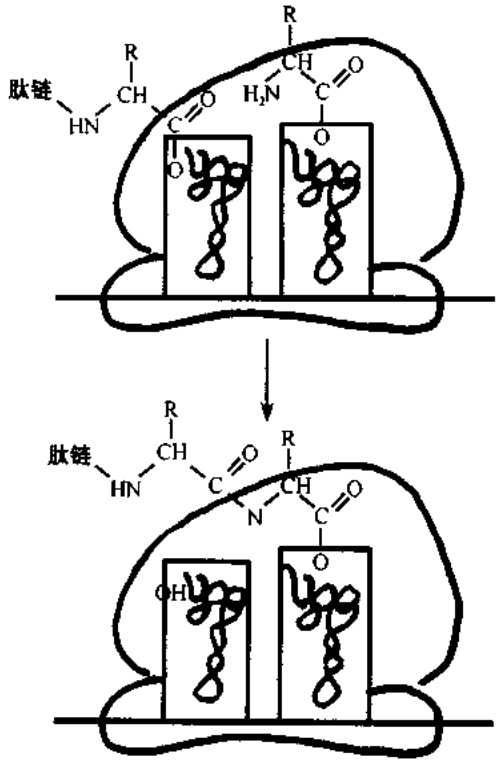


图 15 - 23　肽酰 - tRNA 和 aa-tRNA 之间肽键的形成。（仿 B.Lewin: Genes Ⅵ, 1997, Fig. 8 - 21）

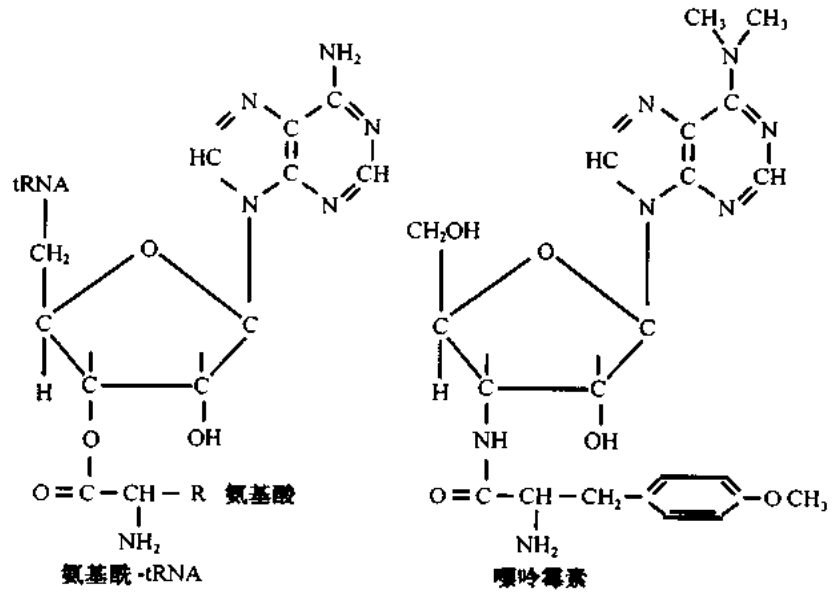


图 15 - 24　嘌呤霉素的结构和 aa-tRNA 的比较。（仿 B.Lewin: Genes Ⅵ, 1997, Fig. 8 - 22）

素有一个 N 取代了将氨基酸和 tRNA 连接起来的"O"。核糖体把它当作进入的 aa-tRNA 来处理，附着在肽酰－tRNA 上的多肽转移到嘌呤霉素的 NH_2 基上。

由于嘌呤霉素的一半不能锚定在核糖体的 A 位点上，因此肽酰－嘌呤霉素从核糖体中被释放出来。蛋白质合成发生早熟终止，抗菌素在其中起到了关键的作用（表 15－5）。

表 15－5 抗生素在蛋白质合成的不同阶段所起的抑制作用

作用阶段	抑制剂	作用对象	抗性突变的作用或特点
起始	春雷霉素	细菌	突变体 16S rRNA 上缺少甲基
	链霉素	细菌	30S 蛋白中的突变
延伸	黄色霉素	细菌	EF-Tu·GDP 不能被释放
	嘌呤霉素	细菌	从肽酰－tRNA 接受肽链，早熟终止
肽的转移	红霉素	细菌	突变存在于 23S rRNA 的修饰
		和粒体	突变存在于 23S rRNA 的顺序
	氯霉素	细菌	突变存在于 50S 蛋白
		线粒体	突变存在于 23S rRNA 的顺序
	放线菌酮	真核细胞质	抑制 60S 上肽的转移
移位	梭链孢酸	细菌	EF-G·GDP 不能被释放
	硫链丝菌肽	细菌	抑制 GTPase 活性

B. Lewin: Genes Ⅳ, 1990, Table 7.6

肽的转移是核糖体大亚基(50S 或 60S)的功能。肽转移酶位于核糖体上肽酰－tRNA 和氨酰－tRNA 的末端彼此靠得很近的位点，从红霉素和氯霉素的作用(表 15－4)可以推测 rRNA 和 50S 亚基蛋白对肽的转移是必要的，实际上此催化活性是核糖体 RNA 和 50S 亚基的一个特点，而不是任何单个的蛋白。

3. 移位反应

氨基酸加到增长的肽链上这一循环是依赖于移位(translocation)来完成的(图 15－25、26)，在图中核糖体延着 mRNA 向前推进一个密码子。移位导致脱酰－tRNA 从 P 位点移出，进入 E 位点使新的肽酰－tRNA 可以进入 P 位点，空着的 A 位点为下一个密码子相应的氨基酰－tRNA 的进入作好准备。

在细菌中空载的 tRNA 通过 E 位点离开核糖体(在真核中它直接进入胞液中)。脱酰－tRNA 和肽酰－tRNA 移位的过程分为两步：首先 tRNA 在 50S 亚基上所处的区域先移入到新的位点，在此阶段 tRNAs 有效地和 50SE/30SP 和 50SP/30SA 位点交错组成的杂合位点相结合。然后 30S 亚基再移动，使密码子和反密码子的配对区处于合适的位置上(图 15－25)。

以前人们普遍认为核糖体上仅有 A 和 P 二位点，即**二位点模型**。1989 年西柏林分子遗传所 K.H.Nierhaus 等从功能上阐明了三位点模型。该模型认为，细菌 tRNA 及 mRNA 相对于核糖体发生移位后，脱酰－tRNA 并未立即从核糖体上解离下来，而是移到了 E 位点（出口位点），当新的氨基酰 tRNA 结合到 A 位时 E 位点的空载 tRNA 才解离下来，此过程涉及到核糖体构象的变化，此变化有利于核糖体对正确氨基酰－tRNA

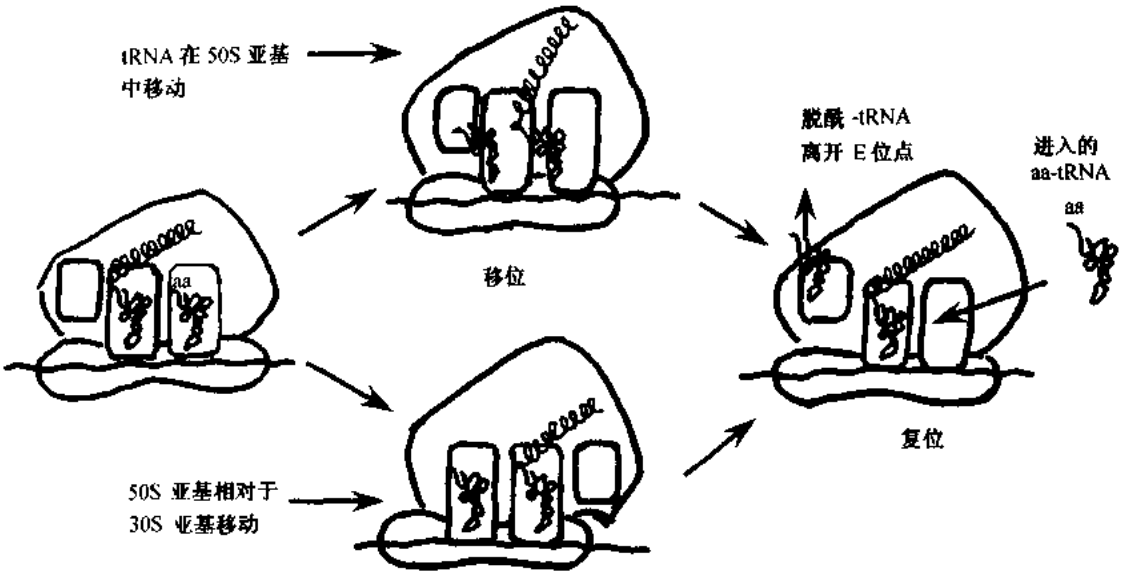


图 15－25　两种移位模型，都涉及到两个步骤，移位和复位。或是 tRNA 本身移动或是 50S 亚基相对于 30S 亚基移动。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8－23）

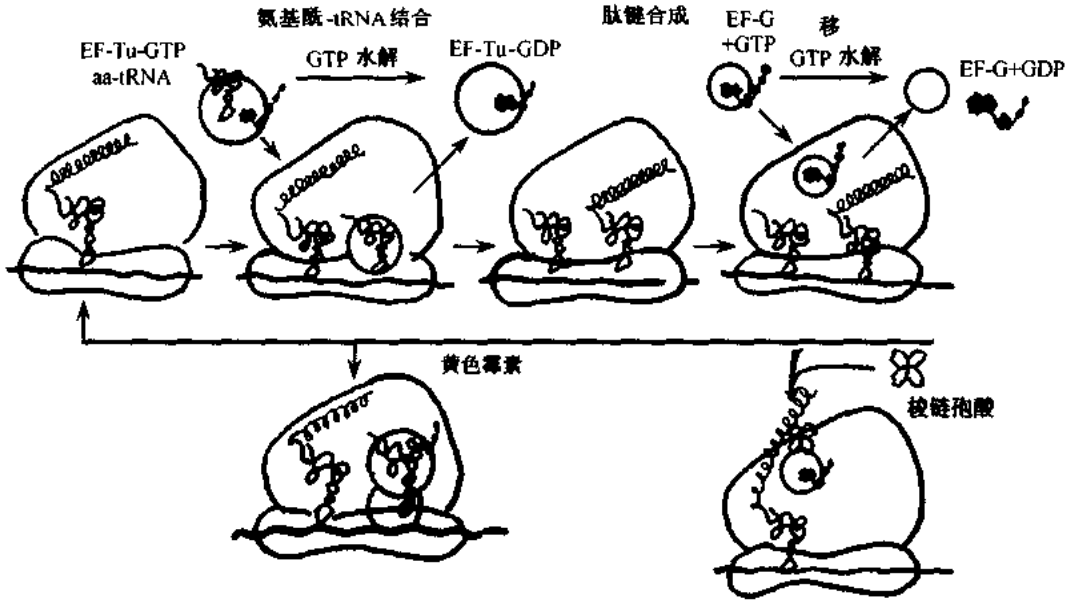


图 15－26　当核糖体接受新的 aa-tRNA 时 EF-Tu 和 EF-G 与核糖体交替结合，形成肽键并移位，梭链孢酸使核糖体转入移位后状态。黄色霉素可导致核糖体拖延，因 EF-Tu-GDP 不能释放。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 8－24）

的识别作用，从而提高了蛋白质合成的精确性。

核糖体具有两种构象状态，一种是移位前状态，A、P 位对 tRNA 有较高的亲和力。即当 E 位占据后 A 位结合减弱；另一种状态是在 A 位点占据后，E 位点减弱，从而空载 tRNA 就 E 位点得以释放。图 15－25 表示了两种移位反应的模型，一种是 tRNA 相对整

个核糖体的移动，这样 tRNA 的氨基酰端在 50S 亚基中移动。移位发生时 tRNA 的反密码子端进一步移动。第二种模型是 50S 亚基先相对于 30S 亚基移动，所以移位分两步进行，到第二步再恢复核糖体的正常结构。在真核生物的核糖体中没有 E 位点，脱酰－tRNA 直接从核糖体上解离下来，释放到细胞质中。

移位需要 GTP 和另一种延伸因子 EF-G(表 15－4,图 15－26)。这种因子是细胞中的一种主要成分。每个细胞中约有 2000 个分子，每个核糖体约有一个拷贝的 EF-G。

EF-G 结合于核糖体上发动移位，当 GTP 水解时它就被释放出来。当 GMP-PCP 取代 GTP 时结合仍然发生。表明鸟苷的存在对于结合是必要的，但它的水解对于移位反应的发生并不是必须的。此因子并不催化 GTP 的水解，此活性是核糖体的功能。

另一涉及蛋白质合成的因子是 4.5S RNA。它是一个 114 碱基组成的，存在于 *E. coli* 中。每个约有 3000 个分子，但其中少量的分子与核糖体结合。4.5S RNA 的突变体生长不良，他们可以被 EF-G 基因的突变所校正，表明 4.5S RNA 和 EF－G 可以某种方式相互作用。4.5S RNA 对于某些分泌蛋白也是必要的，但不是对所有的分泌蛋白。目前尚不清楚它和 EF-G 相互作用的特点。

核糖体不能同时和 EF-Tu 及 EF-G 结合，因此蛋白质合成按图 15－26 的循环进行。EF-Tu 和 EF-G 交替和核糖体结合并从其中释放出来。在 EF-G 结合前，EF-Tu·GDP 必须被释放出来；而在 aa-tRNA·EF-Tu·GTP 结合前也必须释放出 EF-G。甾类抗生素(steroid antibiotic)**梭链孢酸**(frsidic acid)的作用结果发现 EF-G 的释放是必要的。核链孢酸使核糖体停滞在移位后的状态。在它存在之前移位已发生：EF-G 结合在核糖体上，GTP 被水解后，核糖体移动一个密码子。但梭链孢酸使得核糖体·EF-G·GDP 复合体变得稳定，因而 EF-G 和 GDP 就保留在核糖体上而不释放，核糖体就不能结合 aa-tRNA，从而使氨基酸不能进一步加到肽链上。黄色霉素也可使核糖体不能移位，因黄色霉素使 EF-Tu·GDP 不能释放。

现在尚不知道每种延伸因子可以排斥另一种延伸因子的这种能力是通过作用于核糖体整个构象来介导还是对于共同的结合位点采取竞争性结合。这种排斥的作用对于蛋白质合成过程有序地进行甚为必要。

两种因子与核糖体的结合都需要 GTP 的存在。但仅仅后者需要 GTP 的水解。由于它们都能利用 GDP 来维持和核糖体的结合。所以三磷酸的形式可能是这些因子获得活性的构象所必需。这种机制使得两种因子只有在 GTP 的伴随下才能进入核糖体，进而执行它们的功能。GTP 的水解提供了改变核糖体构象的能量。EF-Tu 是 aa-tRNA 进入 A 位点所必需的，EF-G 是核糖体移位所必需的。

在真核生物中相当于 EF-G 的蛋白质因子是 eEF-2，它以相似的方式来执行功能——作为一种依赖于 GTP 水解的移位酶。它的作用也受到梭链孢酸的抑制。一个稳定的 eEF-2 和 GTP 的复合体能被分离出来。此复合体和核糖体结合使 GTP 水解。

eEF-2 对白喉毒素 (diphtheria toxin) 具有敏感性。此毒素用 NAD(nicotinamide adenine dinuclestide)(辅酶Ⅰ)作为辅助因子将半个 ADPR(adenosine diphophate tibosyl；腺嘌呤核苷二磷酸核糖体糖基)转移到 eEF-2 上，此 ADRP-eEF-2 偶联物在蛋白合成中是没有活性的。这个连接物的底物是一种稀有氨基酸——经过修饰的组氨酸。它和 eEF-2 的很多片断是相同的。白喉毒素的致命作用主要是依赖 ADP－核糖基化，这个反应的效率很高：一个单个的毒素分子所修饰的 eEF-2 分子足以杀死细胞。

二、肽链合成的终止

只有61个密码子可以编码氨基酸。另外的3个密码子是终止密码子（termination codon or stop codon），它们结束蛋白质的合成。在发现它们时人们给它们起了别名。UAG叫琥珀（amber）密码子；UAA是赭石（ochre）密码子；UGA有时称为蛋白石（opal）密码子。

这3个密码子的特点原先是通过遗传检测表现出的。存在两种不同类型的点突变：

（1）一种点突变是改变了的密码子对应一个不同的氨基酸，这种突变称为**错义突变**。使得蛋白质中的一种氨基酸被另一种氨基酸所取代；对蛋白质功能的影响取决于突变位点和氨基酸取代的特点。

（2）当点突变产生一个终止密码时，可导致蛋白质合成在突变位点处提前终止。由于在突变细胞中仅产生蛋白质前部的片断，所以这种蛋白质常常是没有功能的。这种突变称为**无义（nonsense）突变**。

有时无义密码子这个词也用于终止密码子。“无义”是个错误的名词，因为这种密码即使在突变基因中中断了蛋白质合成，但仍有含义的。

在已测序的基因中，终止密码子总是紧接在C－端最末一个氨基酸的后面。任何一个三联密码发生无义突变时都足以终止其基因的蛋白质合成。

在细菌的基因中，UAA是最常用的终止密码，UGA又要比UAG重要，虽然阅读UGA存在更多的错误。（错误阅读终止密码就是一个aa-tRNA对它产生错误反应，使蛋白质合成继续进行，直到另一个终止密码子出现。）

没有一个终止密码子具有相应的tRNA。它们的功能完全不同于其他的密码子，而是直接被蛋白质因子所识别。由于这个反应并不依赖于密码子和反密码子的识别，看来没有什么特殊的理由需要三联体顺序。大概这反应了遗传密码的进化。

在*E.coli*中有两种相关蛋白催化终止，它们被称为**释放因子**（release factors（RF）），它们对不同的终止密码子是特异的。RF1识别UAA和UAG；RF2识别UGA和U氨基酸。这两种因子作用于核糖体A位点且需要P位点的肽酰－tRNA存在。释放因子存在的量比起始因子或延伸因子低得多，每个细胞约600个分子，和核糖体的比约为1∶50。可能开始只有一种释放因子，识别所有的终止密码子，而后来它进化成两种因子，具有了密码子的特异性。

RF基因发生突变减少了终止的效率，显然通过增加蛋白质持续合成的能力能通读终止密码子。相反此基因中的任何一个的过量表达将增加在终止密码子处的终止效力。此表明两RF中的任一个因子与错误识别终止密码子的aa-tRNA竞争识别终止密码子。

在真核系统中只有一种释放因子——eEF可识别3种终止密码子。eEF与核糖体的结合需要GTP的，而在细菌中则不需要。可能GTP在终止发生后裂解，随之eEF从核糖体上解离出来。

终止反应涉及到完整的多肽从最后一个tRNA上释放出来，tRNA也从核糖体中排出。这个反应可能的过程是释放因子识别终止位点并与之结合，激活肽基转移酶，水解了P位点上多肽与tRNA之间的键，然后释放了多肽和tRNA。现在尚不清楚核糖体如何从mRNA上解离下来。有可能是RF因子的触动改变核糖体的构象，导致了解离。

16S rRNA参与对终止密码子的识别，其1199－1204位点的顺序为UCAUCA，位于

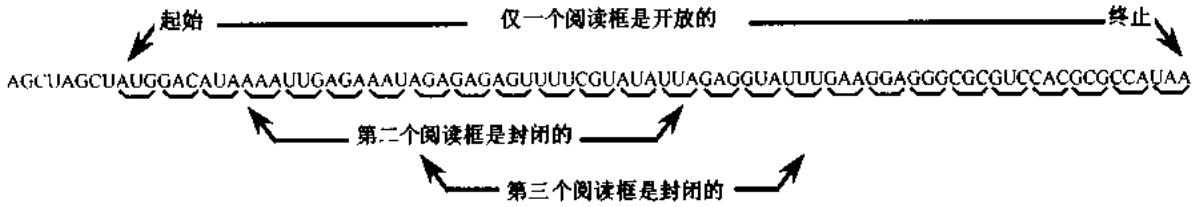


图 15－27 一段顺序的三个阅读框常只有一个开放阅读框，其余 2 个是封闭的。(仿 B. Lewin: Genes Ⅴ, 1994, Fig. 7－27)

二级结构的茎上，茎的另一条链上 1054 位是 C，与 UGA 的识别有关，如缺失这个 C，则可通读 UGA，但对 UAG，UAA 氨基酸无影响。上述顺序的 UCA 若突变成 UUA 也可引起 UGA 的通读。

一个 DNA 顺序可能有 3 种阅读框，但只有一种具有编码的作用，称为**开放阅读框**(open reading frame, ORF)。有的阅读框因终止密码出现频繁故不能生成蛋白，这种阅读框称为**封闭阅读框**(block reading frame)。若一个顺序所有的三个阅读框都是封闭的，则它无编码蛋白的功能。一个翻译成蛋白的顺序有一个阅读框，开始于 AUG 起始密码子，通过一系列有义密码子，直到终止密码子结束。通常 3 个阅读框中总有封闭阅读框的存在(图 15－27)。

当获得了一个未知功能的 DNA 区域的顺序，要通过分析来确定阅读框是开放的还是封闭的。在任何一个 DNA 序列中往往只有一个开放阅读框。ORF 似乎并不是随机存在。如果一个序列不翻译成蛋白质，那么将无选择压力来阻止其中产生无义密码子。这样一个长 ORF 的鉴别，乍看起来这个序列好像是可以翻译的，但也鉴别出无蛋白产物，有时称其为**非识别阅读框**(unidentified reading frame，URF)。

第四节 真核生物蛋白质的越膜运输

一、蛋白质的定位装置

真核细胞是一个高度有序的结构，很明显作为大分子结构成分的蛋白质要被导入特殊的位点，除了细胞中的酶以外，所有的蛋白都要被导入到特殊的目的地。例如位于卵中的蛋白进入到极端分化的细胞中，在某些物种卵母细胞中初始分布不对称的蛋白可能为器官发育所需要。特殊细胞的功能也常常是由特殊蛋白所构成的大分子所决定的。定位是各种真核细胞内部结构所必要的。结构的决定是取决于各种蛋白正确的定位。

我们可以根据合成的位置和最终到达的位置对蛋白分类。

(1)“可溶性”蛋白不定位于任何特殊细胞器中，他们在细胞液中合成且保留在那儿，作为单个的催化中心，作用于胞液中可溶性的代谢物。

(2) 由胞液中蛋白构成的大分子结构可能定位于胞质中的特殊位点，例如中心粒要和有丝分裂纺锤丝的极区结合。

(3)核蛋白必需从在细胞合成的地方穿过核膜转移到核中。很多蛋白是染色质的成分，其他的构成核纤层蛋白，核骨架或外膜。还有的是调控蛋白，担负着转录调节的

使命。

(4)细胞器中的有些蛋白是在胞液中合成并转移(穿过)细胞器膜,例如植物中的线粒体、叶绿体。线粒体和叶绿体本身也具有合成蛋白质的能力,在细胞器中还合成其他一些蛋白,这些蛋白就留在该细胞器中。

(5)细胞质中含有各种有膜的小体,包括内质网(ER),高尔基体,胞内体和溶酶体。这些都是属网状内皮系统或液胞。要定位在这些成分中的蛋白质在开始合成时插入 ER 的膜上,然后通过高尔基体的易位系统导入它们的特殊靶位点。

从细胞中分泌的蛋白必须穿过质膜排出外表面。它们以相同的方法,即蛋白质和网状内皮系统结合开始它们的合成,然后要越过整个的系统,而不是停留在其中的某些地方。

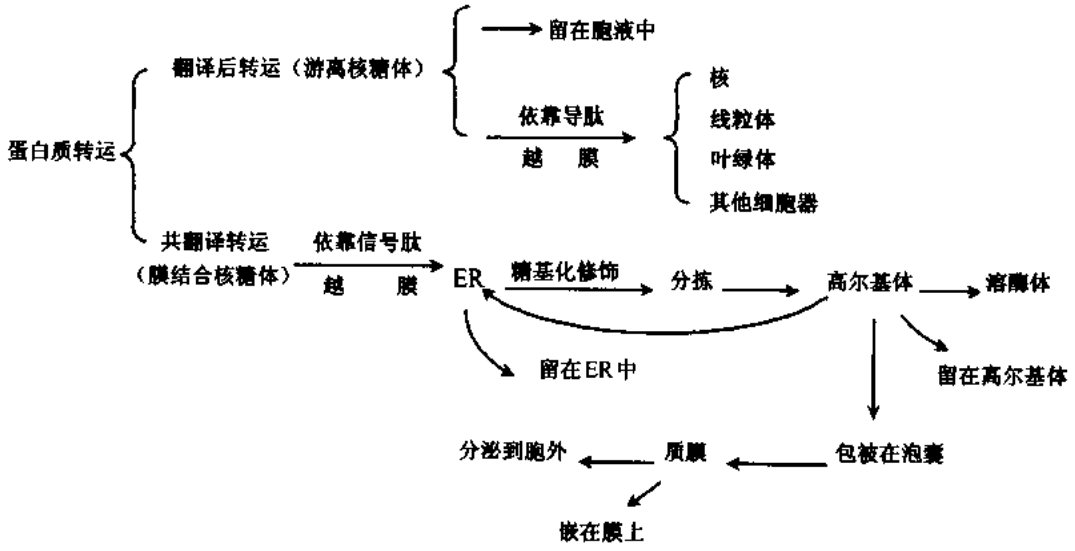


图 15-28　蛋白质合成后的转运,定位及分泌。

图 15-28 表示细胞中蛋白质合成的位置以及合成后蛋白质的定位及转运。除了可溶性的胞液蛋白以外,合成的蛋白开始以一定的途径结合到适当的位置上。**蛋白分检**(protein sorting)或**蛋白质的运输**(trafficking)依赖于蛋白质和细胞膜系统相互作用的能力。或通过蛋白质顺序的特点,或通过在其上共价基团,使一种蛋白质可被特殊膜上的受体识别。这种蛋白穿越过膜而进入某种细胞器的过程就被称为**蛋白转运**(protein translocation),根据其目的地的不同,转运蛋白可能需要不同的分拣信号和转运信号使其输入到细胞器中或者穿越过膜。

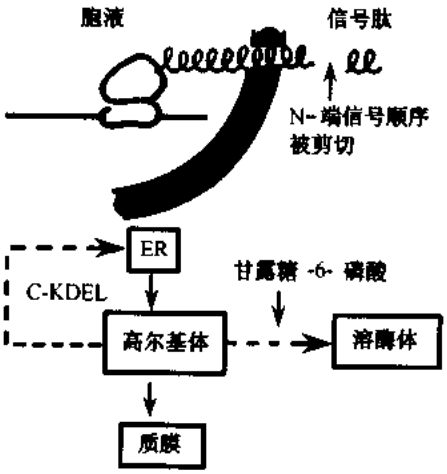


图 15-29　膜结合核糖体的信号肽和转运。
(仿 B.Lewin: Genes V,1994, Fig.11-3)

二、导肽

蛋白质的合成开始于多核糖体,而多核糖体可分为两类:**游离的和膜结合的**(图 15-29)。它们承担不同类

型蛋白质的合成。

游离的核糖体负责合成释放到胞液中的蛋白。蛋白运转的目的地取决于这些蛋白是否具有特殊的定位信号。若没有任何特殊的信号，这种蛋白只有以可溶的形式留在胞液中。移向不同目的地的蛋白具有各种不同的信号。这种信号以一种短的"模体"形式存在。信号的功能是在蛋白质合成完成后才发挥作用的，所以这过程称为**翻译后转运**（post-translation translocation）表 15－6 表示由游离核糖体合成的蛋白所用的信号顺

表 15－6　导肽引导蛋白进入的组织及其特点

组织	信号的位置	类型	信号顺序的长度
线粒体	N 端	氨基酸	12～30aa
叶绿体	N 端	氨基酸	约 25aa
核	中部	碱性氨基酸	7－9aa
过氧化物酶体	C 端	SKL(Ser Lys Leu)	3aa

序，通常被导入线粒体和叶绿体，在蛋白 N 端有长约 25aa 的特殊顺序，此顺序可被细胞器被膜上的受体所识别，且在越膜后常被切断。进入核中的蛋白，这段氨基酸顺序更短。这些**核定位信号**(nuclear lacalization signals)可穿越核孔。转运到过氧化物酶体的蛋白是通过一个非常短的 C－端顺序。

膜结合多核糖体(membrane-bound polysomes)中，当多肽一开始合成时，就立即被导入 ER 中，这一过程就称作为共翻译转运(co-translational translocation)。信号肽常出现在

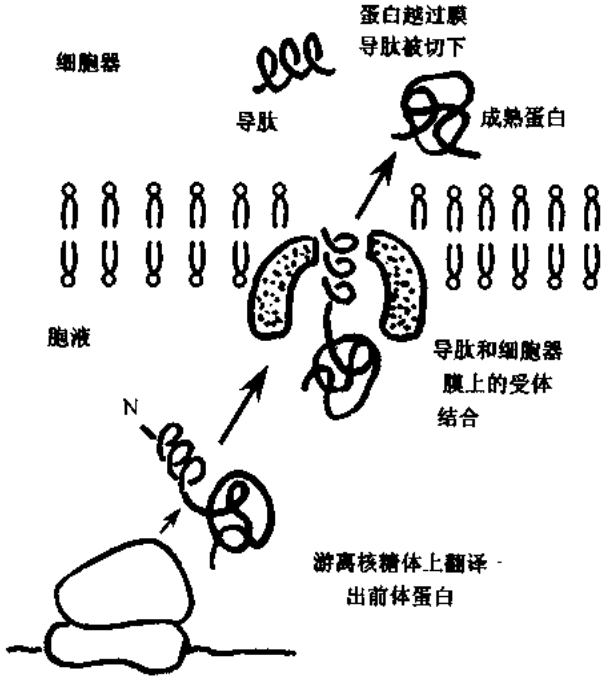


图 15－30　通过翻译后加工导肽使蛋白向线粒体或叶绿体的表面转运

N 端，但有时也有在蛋白质的中部或 C 端，取决于此蛋白最终转运到何处而定（图 15－30）。分拣的原则要看蛋白定位在膜系统的什么部位。通常信号是由一个短的aa 顺序所提供，但也有其他类型的信息。剪切发生在通过 ER 时，然后进入高尔基体进一步转运到质膜上。留在 ER 中的蛋白 C 端具有4 肽(C-KDEL; Lys-Asp-Glu-Leu)，实际上C-KDEL

是从高尔基体进入 ER 的一个信号。将蛋白引入到溶酶体的信号是一个共价修饰：加上特殊的糖基甘露糖 - 6 - 磷酸。

无论是翻译后转运还是共转运其共同特点是转运蛋白的 N 端有一段特殊序列，这段序列在蛋白转运时要被剪切掉。N 端序列由一个**前导肽**(leader)组成。在成熟的序列上不存在这部分，表明它是一段蛋白质上的附加区域，作为一种稳定的前体存在。

蛋白质通过 N 端的前导序列和膜结合，用不同等级的信号来寻找其最后的归宿。前导肽本身将蛋白导向内膜，这种与膜相互作用的保守序列负责蛋白质穿越过膜进入它的特异位置。但蛋白质的其他序列也可以使蛋白质留在膜上，取决于它所采用的构象。位于膜中的蛋白称为**整合膜蛋白**(integral membrane protein)。

能进入膜内的能力看来是前导序列的功能，但种一蛋白能留在膜中主要依赖使蛋白产生折叠的信息。蛋白保持去折叠是因为如果成熟的蛋白体积太大就不能有效通过一种孔道。另一个原因是当蛋白质通过疏水性的脂双层 (hydrophobic lipid bilayer) 时，需要暴露亲水的基团，一旦穿过膜以后蛋白就重新折叠成成熟的构象。

三、导肽的结构和功能

翻译后膜的插入依赖于前导序列。

线粒体和叶绿体都能合成其本身的核酸和某些蛋白。线粒体只合成约 10 种细胞器蛋白，叶绿体合成 20% 的细胞器蛋白。大部分细胞器蛋白是在细胞液中合成的，和通过游离的核糖合成胞液的蛋白方式相同，合成后再转运到细胞器中。

进入线粒体或叶绿体的蛋白要通过翻译后加工，初始合成的前体要比成熟的蛋白长 12 ~ 20aa，前导肽负责细胞器外膜的初始识别，图 15 - 30 表明导肽起始了前体蛋白和细胞器膜的相互作用。导肽穿越过膜后被细胞器的蛋白酶切下。转运继续进行整个

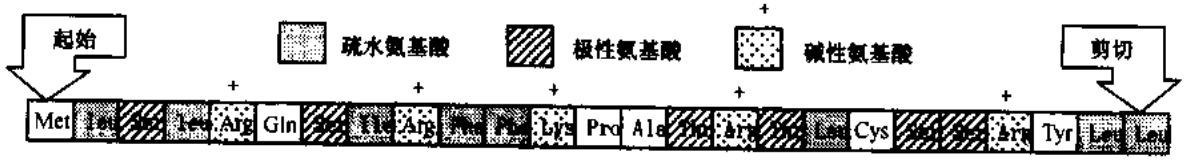


图 15 - 31　酵母细胞色素 c 氧化酶亚基 Ⅳ 的导肽含有 25 个中性和碱性的氨基酸，前面的确 2 个氨基酸可有效地将任何附加的多肽转移到线粒体中。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 10 - 10)

的蛋白都穿越过膜，或直到中部特定序列出现导致转运的停止。输入到线粒体或叶绿体的蛋白的导肽有很少的同源性。

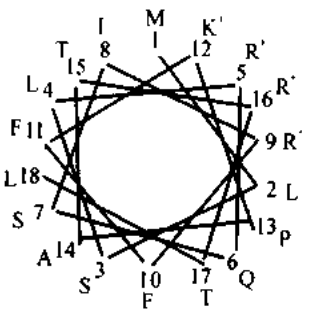


图 15 - 32　双亲螺旋轮。

导肽通常是疏水的。由非电负性氨基酸构成，中间夹有碱性氨基酸，而没有酸性氨基酸，羟基氨基酸含量高 (特别是 Ser)，易形成双亲螺旋轮 (图 15 - 32)。不同的前导序列缺乏同源性，这意味着和识别有关的信号不是一级结构，而是二级三级结构，或这一区域产生的特点。有一种可能是此前导肽形成一种**双亲螺旋** (amphiphilic helix)，即带电荷的基团在螺旋的一侧，不带电荷的在另一侧。

前导肽含有作为细胞器蛋白定位所需的所有

信息。导肽的这种能力可以通过构建人工蛋白来检验。具体方法是构建融合基因(fusion gere)然后翻译成融合蛋白。将来自细胞器的导肽和位于胞液中的蛋白连接起来。

各种导肽通过这种实验表现出具有独立引导附着序列进入线粒体或叶绿体靶位点的功能。如将图 15－32 所示的导肽加到小鼠胞液蛋白二氢叶酸还原酶 (DHFR, dihydrofllate reductase)上，此 DHFR 就可变成定位于线粒体的蛋白。

导肽和转运蛋白存在着独立折叠的结构域。如不考虑附加上的序列，导肽必须能折叠成一个适当的结构供细胞器表面受体识别。附加序列对识别不起作用。

越膜运输需要 ATP 的水解，蛋白磷酸化和转运直接相关，它需要 ATP 为转运装置的成分提供能量，或者为转运时多肽的去折叠所需。看来越膜转运，导肽与膜的结合早期阶段并不需要 ATP 水解，而是进入膜以后或转运的本身需要 ATP。

是什么为越膜提供了力量呢？进入线粒体（及细菌的输出）需要一种电化学的电势(electrochemical potential)使导肽的 N－端穿过内膜(inner membrane)进行转运。蛋白质其余部分的转运并不需膜电位，这就意味着导肽结构就足以越过脂双层所构成的障碍。

是什么限制了转运的亲水蛋白通过疏水性的膜呢？氨甲喋呤(Methotrexate)的作用解释了这一问题。它可作 DHFR 的标识，并阻碍线粒体导肽融合的 DHFR 转运到线粒体中与靶结合，氨甲喋呤在 DHFR 越膜时阻碍了此酶形成折去叠构象。因此转运蛋白的序列与靶目标的识别无关，但为了使它随着导肽穿过膜就需要形成去折叠的构象。

穿越线粒体膜需要一系列的折叠和去折叠。当完成合成后，新合成的蛋白被释放到胞液中后就形成了折叠构象。而穿越线粒体膜时又需要去折叠的构象。蛋白越膜以后以去折叠的状态存在，但需进一步折叠成成熟的构象。最后的折叠需要一种附属蛋白——**分子伴侣**(chaperonin)的存在。

线粒体输入装置的很多成分已被鉴别出来，在输入的第一阶段是与外膜 (outer membrane)上的受体结合，蛋白再进入输入孔道。各种受体对于输入蛋白存着特异性。有的蛋白只有一种受体，有的有多种受体。

当已输入的蛋白锚定在基质上时，它和基质蛋白 hsp(heat shock protein) 70 结合，这是一种和其他应激(stress)蛋白 hsp70 有关的分子伴侣。应激蛋白或热激蛋白都是因应激而合成的分子伴侣，其作用是和不适当折叠的蛋白相结合。很可能当 hsp70 从内膜露出时对于蛋白的去折叠构象的形成是有帮助的，它可帮着把蛋白拉过孔道。

在线粒体基质中应激蛋白 hsp60 等分子伴侣。hsp60 是以一种大的寡聚体式存在(14 个亚基，60kDa)。它和细菌的分子伴侣 GroEL 同源。当 hsp60 和输入蛋白结合时，它维持这些蛋白的松散折叠的形式。通过 ATP 的裂解它可被释放出来。此 hsp60 寡聚体可以提供一个基质蛋白上的支架将输入蛋白装配成成熟型。输入蛋白亚基连接形成寡聚复合体需要与 hsp60 结合，一种输入蛋白在形成适当构象的过程中可以由 hsp70 传递给 hsp60。

四、导肽在线粒体叶绿体中决定蛋白定位

1. 线粒体的蛋白转运

线粒体是由含有二层膜的外膜所包被。蛋白质转运到线粒体中可能被定位在外膜，内外膜的间隔，内膜或基质中。若一种蛋白是膜的一种成分，那么它可被定向在膜

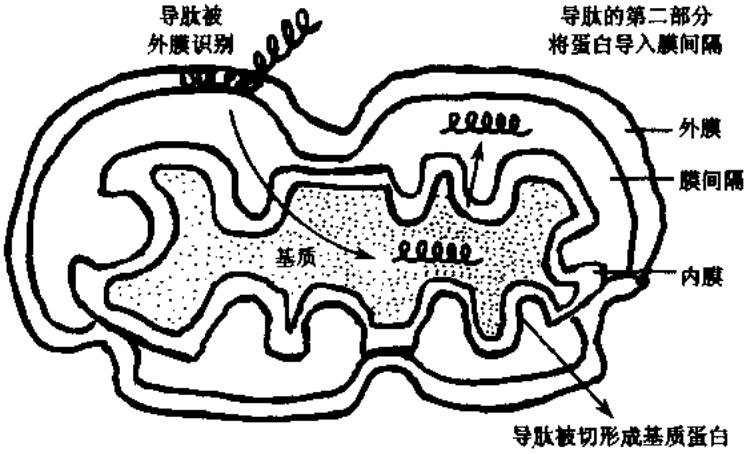


图 15 - 33　细胞器蛋白的不同定位。(参考 B.Lewin: Genes Ⅵ, 1997, Fig. 10 - 11)

的外侧或内侧表面。

是什么指导线粒体蛋白进入其适当的位置呢？某些蛋白质穿越两层膜进入基质的，此特点是由导肽 N 端所赋予的。一个蛋白它定位在膜间质中或在内膜需要另一个附加的信号，此信号对于蛋白在细胞器中的定位是特异的。图 15 - 33 表示导肽上不同的信号分别负责不同层次的定位功能。导肽的第一部分是负责将蛋白的目标定为细胞器。第二部分是负责其目的地不到基质。导肽的这二个部分可被连续切除。

例如细胞色素 c1，它结合在内膜和膜间质的表面，其导肽含 61aa 且可分为不同的功能区。完整的导肽可以转运一个附着蛋白(例如成熟的 DHFR)进入到膜间隔中，但前 32aa 甚至此区域 N 端的一半就可将 DHFR 转运到基质中，因此导肽的第一部分 (N 端的 32aa)构成了**基质导向信号**(matrix-targeting signal)。

为什么完整的导肽将蛋白导向膜间质呢？是由于在基质导向信号区域的后面邻接着 19aa 构成的另一功能区，它将蛋白定位在内膜或在膜的间隔区。这一区域被称为**膜导向信号**(membrane-targeting sigmal)。

基质导向信号的剪切是单独加工的，这一加工对于定位在基质中的蛋白来说是必要的。对留在膜间隔中的蛋白，这一信号也必须切去。此信号剪掉后膜导向信号就将蛋白导入外膜膜间隔或内膜的合适位置。然后再依次剪切。

单个的蛋白酶涉及到基质导向信号的剪切，不论蛋白最终目标。这种蛋白酶是水溶性的，Mg^{2+}依赖性酶，它位于基质中。因此 N 端的顺序必需要达到基质，即使是其最终的目标在膜间质。

N 端的基质导向信号的功能可能在所有线粒体蛋白中都是相同的，它被外膜受体识别，导致蛋白穿越过二层膜。因此所有进入线粒体的蛋白转运开始过程都是相同的，无论它的最终定位在何处。

导肽的特点决定了随后的转运和定位。位于基质中的蛋白除基质导向信号外没有其他的信号 (图 15 - 34)。如有膜导向信号需要切除掉基质导向信号后才能显示其功能。切除后导肽保留的部分(新的 N 端)可使蛋白质到达其最终目的地。

蛋白质用什么方法穿过基质从一层膜到另一层膜呢？转运发生的地方有可能是与外膜相连或融合的地方。当前体主要的部分露在外膜表面的胞液中时，就能捕获到基质酶剪切导肽的中间过程。表明一个蛋白在穿越时跨在两个膜上。导肽亲水的特点

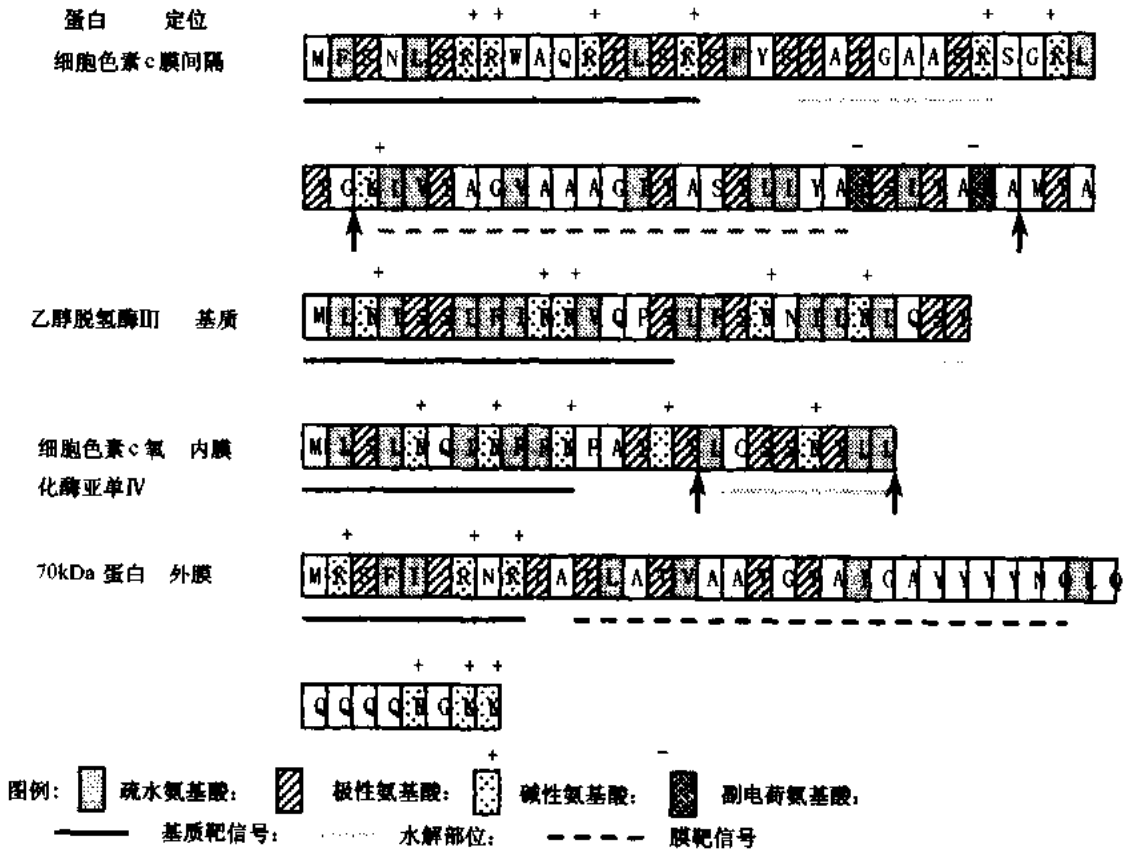


图 15－34 导肽的氨基酸顺序及不同的功能区。

表明输入蛋白要穿越一个由脂双层形成的亲水孔道。

膜靶信号的特点是有争论的。一种模型认为完整的蛋白先进入基质，然后膜靶信号再将它导向并通过内膜。另一种模型提出，线粒体蛋白依靠导肽不同分别牵引到基质，内外膜间隔的外膜。

含有两种信号的导肽其两个部分构成不同，图 15－34 表明 N 端的 35aa 与其他的导肽相似，非负电性氨基酸含量很高。中间间隔着碱性 aa，邻接的 19 个连续的非负电性氨基酸，其长度足以跨越脂双层。这一序列和涉及转运到 ER 膜中的信号肽的序列相似。

可剪切的导肽是，使蛋白定位于细胞器中唯一的信息形式。有的线粒体蛋白是以其成熟的形式被识别的。它们有一段膜序位于 N 端或中部，这段顺序无需剪切就可发动越膜。在任何情况下剪切都不和细胞器的识别机制相关。因此剪切位点的突变也并不阻止蛋白的输入。

2. 叶绿体蛋白的转运

叶绿体蛋白需要控制穿越二层膜。图 15－35 表明各种定位的叶绿体蛋白。这些蛋白穿过外被的外膜和内膜进入基质（stroma），此过程与蛋白质越膜转运到线粒体的基质(matrix)的情况相同。但有的蛋白还要进一步穿越类囊膜(thylakoid membrane)进入腔(lumen)中。定位于类囊膜或腔中的蛋白转运途中必须经过基质。

叶绿体的导向信号和线粒体的相似。导肽由 50aa 组成，N 端的一半是叶绿体外被识别所需。在穿越外被时或穿越后在 20～25 位点之间发生剪切。定位在类囊或腔中的

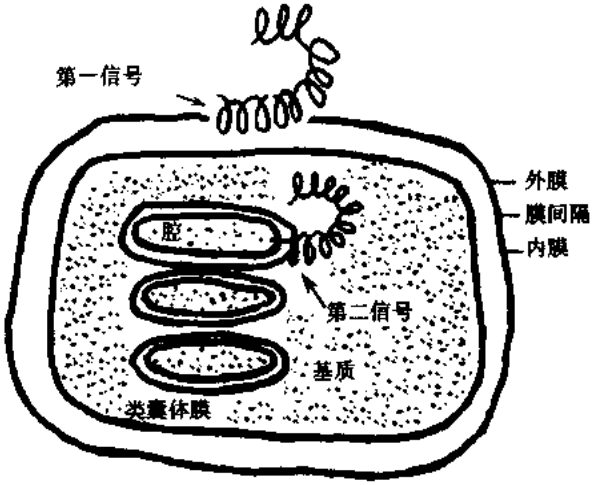


图 15 - 35　各种不同定位的叶绿体蛋白。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 10 - 13)

蛋白由 N 端余下的一半导肽指导类囊膜的识别。

控制蛋白质转运到线粒体和叶绿体中的一般原理是定位于细胞器基质蛋白,其导肽的 N 端部分及其中附加序列是蛋白质定位于外膜,膜间隔和内膜所需要的。

无论是转运到线粒体还是叶绿体的蛋白,其导肽都具有相同的特点。绿色藻类的叶绿体导肽构成的融合蛋白能在酵母中将附加 DHFR 顺序转运到线粒体中,但其效率要低于真正的线粒体转运顺序。

在植物细胞中合成的线粒体蛋白和叶绿体蛋白有什么区别呢?导肽的相似点表明细胞器的蛋白转运的机制相同。对于一个合适的导肽来说,各种细胞器的受体相对的转运效率是不同的。但现在还不清楚原因何在。

3. 过氧化物酶体的蛋白转运

导肽另一个例子是位于 C 端的氨基酸顺序。过氧化物酶体(peroxisomes)是一种由单层膜包被的小体,它含有与氧利用有关的酶。它们通过从底物中除去氧原子将氧变成为过氧化氢(hydrogen peroxide)。然后催化用过氧化氢来氧化其他各种底物。在过氧化物酶体中的所有酶都是从胞液中输入的。就像转运到核中一样,转运到过氧化物酶体的过程是依赖一短的顺序在翻译后进行。各种过氧化物酶体的酶都有 C 端顺序 SKL (Ser-Lys-Leu)。加在 C 端这种很短的顺序除作为一种输运的工具,可能还有其他的功能。

五、信号肽的结构和功能

各种结合在膜上或越膜的蛋白其特点是利用导肽上的各种信息来到达目的地。然而在一个细胞器外被翻译后再转运的导肽与协同翻译进入分泌途径的导肽的作用是不同的,后者常称为**信号肽**。

留在 ER 中、高尔基体中、质膜中或分泌到细胞外的蛋白它们与膜结合有一个共同特点。合成这些蛋白的核糖体与 ER 结合,这样新生蛋白能以共翻译的形式转运到 ER 中。

ER 可以分成两种类型:①膜结合多体的称为**糙面内质网**(rough ER, RER),为扁囊

网；②未结合多体的称为**光面内质网**(smooth ER)，为小管网。ER 在细胞中特别突出，分泌蛋白都在 ER 上合成。

在 RER 上合成的蛋白质从核糖体直接越过膜进入 ER，然后，蛋白质从 ER 膜再转运到高尔基体。导向它们的最终目的地。如溶酶体，或分泌胞，或质膜，如免疫球蛋白和多肽激素都是通过此途径分泌到细胞外。

1. 信号假设的建立

1972 年 Milstein 等发现免疫球蛋白 IgG 轻链的前体要比成熟的 IgG 在 N 端多 20 氨基酸。他们推测这 20 个氨基酸可能和通过 ER 进而分泌有关。美国 Bloble 实验室完成三项重要的实验支持了以上推测：①IgG 的 mRNA 在无细胞系统中，以游离核糖体体外合成时产生的蛋白是 IgG 的前体；若在无细胞系统中加入狗胰细胞的 RER，就能产生 IgG 成熟蛋白。成熟的 IgG 轻链蛋白和前体蛋白相差的 20 个氨基酸是疏水性很强的氨基酸；②加入蛋白水解酶不能使正在合成的 IgG 水解，而同时加入去垢剂就可以使其水解。由于蛋白酶只能作用于游离的蛋白而不能作用与膜结合的蛋白，所以表明 IgG 合成可能和 RER 的膜是结合的，而用去垢剂可将其和膜分离才得以水解；③用去垢剂处理骨髓瘤后所获得的多核糖体与膜分离，然后在离体的条件下继续进行新生肽的合成。经短时温育得到的是成熟的 IgG，而长时温育得到的是前体 IgG，此表明 mRNA 5′端核糖体上合成的新生肽尚未来得及加工，而在 3′端核糖体上合成的新生肽在核糖体未分离前已部分进入 RER，经过了加工，切除了 N 端的部分。

在以上实验的基础上 Bloble 和 Dobberstin(1975) 提出了**信号假说** (signal hypothesis)，认为分泌蛋白 N 端有一段信号肽，当新生肽长约 50～70 氨基酸后，信号肽从核糖体的大亚基中露出，立即被 RER 膜上的受体识别并与之结合。在信号肽越膜进入 RER 内腔后被信号肽酶水解。正在合成的新生肽随着信号肽通过 RER 膜上的蛋白孔道穿过脂双层进入 RER 腔内。这一假设经过多年的继续研究又有了新的拓展，但基本观点仍是正确的。Bloble 因这项成就而荣获了 1999 年度诺贝尔生理学和医学奖。

2. 信号肽的结构和功能

对哺乳动物核糖体和 ER 膜的研究建立了膜插入机制模型。这些系统能将新生蛋白(nascent protein)包被起来进入膜。当将翻译后的前体蛋白分离出来时这个系统就不起作用。这表明合成开始时必须要和膜结合。

分泌蛋白的 N 端通常由可被剪切的 15～30aa 的前导顺序组成，此顺序称为**信号肽** (signal sequence)，在 N 端或靠近 N 端处有 2～3 个极性 aa，而在信号肽的中部都具有惟一的疏水核心或者由很多的疏水 aa 构成。信号肽中没有其他保守顺序也没有酸性基团(图 15－36)。

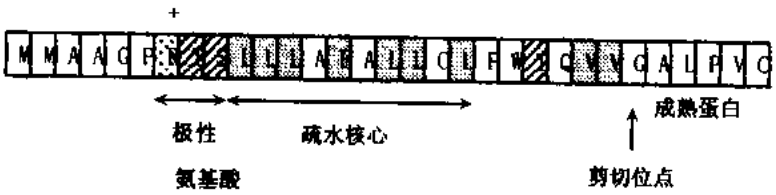


图 15－36 信号肽的结构。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 10－14)

和导肽相似，信号肽也可将任何附加的多肽转运进靶膜。例如将信号肽加在珠蛋白的 N 端，就可使它不再留在胞液中，而是穿过膜而分泌到胞外。

信号肽可使正在翻译的核糖体附着到 RER 膜上。核糖体是通过信号肽的功能而附着并合成分泌蛋白的。因此游离的核糖体和膜结合核糖体之间本身并无差异。

信号肽作为一种附着到 ER 膜上的信号识别，可能是通过开始合成出的 N 端头几个氨基酸的疏水功能。然后蛋白链插进膜中，信号肽被埋在膜中的一种蛋白酶所剪切，这时核糖体已完成翻译，蛋白已延着导肽途经穿过膜。

盐冲洗过的膜不能启动核糖体的结合，取消盐洗，它的能力又可以恢复。盐洗可灭活的活性成分叫做**信号识别蛋白**(signal recognition particle ，SRP)。它是一个宽 5～6nm，长 23～24nm 长条状的结构，且能分离出 11S RNP 复合体，含有 6 种蛋白(总分子量为 240kDa)和一个小的 7S RNA(305 碱基，100kDa)。此 7S RNA 提供此蛋白的结构骨架，没有这个骨架单个的蛋白不能装配。

SRP 有三个重要的功能：①它能和新生的分泌蛋白的信号肽相结合；②还能和位于膜上的蛋白受体相结合；③延伸制动。

有功能的 SRP 可由一种 7S RNA 和其他一些蛋白组装而成。像其他转运和越膜蛋白一样，SRP 普遍存在于真核生物中。

SRP 和 SRP 受体二者的催化功能是将带有新生肽的核糖体转移到膜上。第一步是

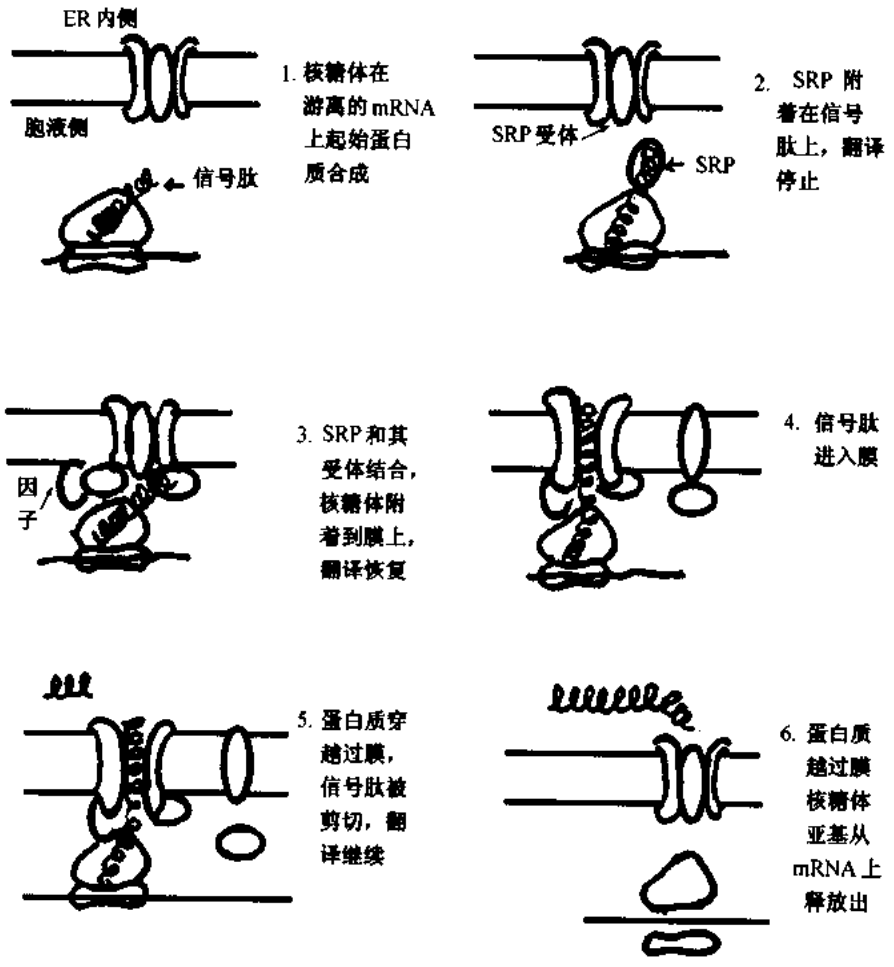


图 15－37　信号肽和 SRP 在蛋白越膜中的作用。(仿 B.Lewin：Genes Ⅵ，1997，Fig. 10－19)

信号肽被 SRP 识别。然后 SRP 和其受体结合，核糖体结合到膜上，图 15－37 表示膜蛋白翻译的阶段。

SRP 受体在蛋白质转运中的作用是短暂的。当 SRP 和信号肽结合时，它阻止了翻译，蛋白合成停止。此是在新合成的多肽链长 70aa 左右时发生的。（这样 25～30 残基的信号肽伸在核糖体外面，相邻的约 40 个 aa 仍在核糖体中）。

当 SRP 与其受体结合时，SRP 释放出信号肽，然后核糖体和膜上的其他成分（尚未鉴别出）结合，此时翻译得到恢复。当核糖体被传递到膜上时，SRP 及其受体的作用已完成了，又进入新的循环，发动另一些新生肽和膜的结合。

SRP 7S RNA 可分成两部分：5′端的 100 个碱基和 3′端的 45 个碱基，这一段和 Alu RNA 顺序密切相关，因此定义为 **Alu 结构域**（Alu domoin）。RNA 余下的部分由 **S RNA 功能域**（S RNA domain）构成。

SRP 的结构如图 15－38 所示，RNP 不同的部位对于靶蛋白具有相应三种功能。在体外用 SRP 的识别来研究每部分的功能。5kDa 的蛋白只有一个分子，它不直接和 RNA 结合。而是和 19kDa 的蛋白结合，19kDa 蛋白与 RNA 的两个末端结合。P54 用来识别信号肽；P68/72 双体结合于 RNA 的中心区域，它是用来识别 SRP 的受体及蛋白的越膜转运。P9/14 二聚体结合在此 RNA 分子另一端的附近。它负责延伸制动。

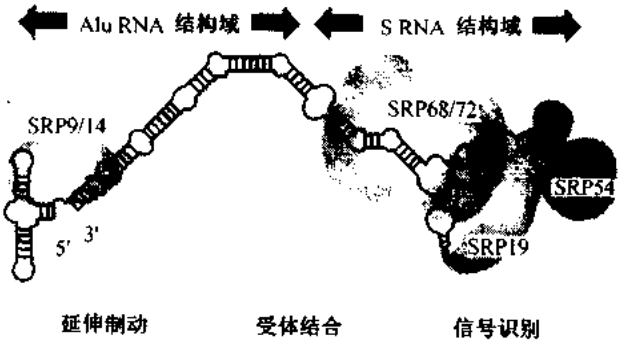


图 15－38　SRP 的结构和功能。（引自 B. Lewin: Genes Ⅵ, 1997, Fig. 10－20）

SRP 的受体是含有 72kDa 和 30kDa 两个亚基的二聚体。大亚基的 N 端锚定在 ER 中，蛋白的大部分伸展在胞液中，蛋白此区域的大部分顺序与核结合蛋白相似，带有很多正电荷的 aa，表明 SRP 受体识别 SRP 中的 7S RNA，SRP 受体小亚基的功能现在还不清楚（图 15－39）。

- SRP
  - 形状——长 23～24nm，宽 5～6nm，11S RNP 复合体，240kDa
  - 组成
    - 7S RNA，305b，100kDa，作为 SRP 的骨架
      - Alu 结构域
        - 5′100nt
        - 3′45nt
      - S 结构域
    - 蛋白，6 种单个无活性
      - P 9/14——二聚体，阻止肽的延伸（干扰进入的 aa－tRNA 和肽酰转移酶）
      - P 68/72——二聚体，识别 SRP 受体，越膜转运，位于 RNA 的中心区
      - P 54——单体，识别信号肽，不直接和 RNA 结合，而和 P19 结合
      - P 19——单体，和 RNA 的两个末端结合，和 P54 结合
- SRP 受本
  - α 亚基——大亚基，72kDa
    - N 端：12K，嵌在膜内疏水部分
    - C 端：60K，暴露于胞液的亲水部分，带有很多正电荷氨基
  - β 亚基——小亚基，30kDa，识别 7SRNA
- 信号肽酶——6 种蛋白组成
  - 其中一种蛋白具有酶的活性
  - 其他 5 种蛋白起修饰作用和构成一定的结构

图 15－39　与转运有关蛋白的结构和功能。

为什么协同翻译的蛋白进入内皮网状系统，而翻译后转运的蛋白就要进入线粒体和叶绿体呢？还没有充分的证据来明确回答此问题，且在酵母中输入 ER 的蛋白却发生翻译后转运，这使问题更为突出。两种过程对能量的需求是不同的。转运到线粒体或叶绿体中需要一种电位差，而进入 ER 时需 ATP。这并不意味着蛋白系统的作用涉及到能量的提供。核糖体的存在对于在协同翻译转运中维持蛋白进入膜中时的正确几何形状是需要的。

一种观点认为越膜转运涉及到蛋白构象的控制。若蛋白的顺序能被释放到胞质中的话，它产生的构象取决于含水的环境。以这种构象它可能不会穿过膜。SRP 的能力是在核糖体和膜接触前抑制翻译，阻止蛋白释放在含水的环境中。然而以这种方法控制构象也仅是在信号肽位于 N 端的情况下，在其他情况下是不行的。因此越膜运输还可能涉其到与转运蛋白结合，直接影响到他们结构的一些因子。

信号肽酶（signal peptidase）（在体外鉴别的）是由 6 种蛋白组成的复合体。实际上只有其中的一种蛋白具有酶的活性，其他的蛋白可能起到修饰作用或者与形成一定的结构有关，如与在膜上的定位或形成膜上的通道有关，它们的量约和结合核糖体的量相等。信号肽酶位于 ER 膜的内表面上。表明信号肽在被切割前必须穿过膜。

膜上核糖体受体又称为**多肽转运装置**（translocation machinery）或**核糖体亲和蛋白**（ribophorin）。可能由于核糖体受体和核糖体接触后，在膜上聚集而形成孔道，使信号肽及其相连的新生肽得以通过。此时 SRP 与其受体分离恢复游离状。翻译和转运完成后，核糖体大、小亚基相互解离，核糖体也发生解聚，通道消失，ER 的膜也恢复完整的脂双层。

六、锚顺序的结构与功能

亲水的蛋白怎样越过疏水膜呢？其 N 端的信号肽实际起到的是引路的作用。人们推测一旦脂双层的障碍被信号肽的进入打破以后，肽就可以连续地穿过膜直到肽被转运到别的地方为止。转运的起始是靠信号肽而不受附加顺序的支配。

越过 ER 膜进行转运的蛋白可以通过变性剂来抽取，但用同样的变性剂却不能抽提作为膜成分的蛋白。图 15－40 表示转运模型。信号肽或在特异位点插入脂双层，或者在膜蛋白的含水孔道中通过。此孔道可以借助于离子通过的能力来检测（在电传导

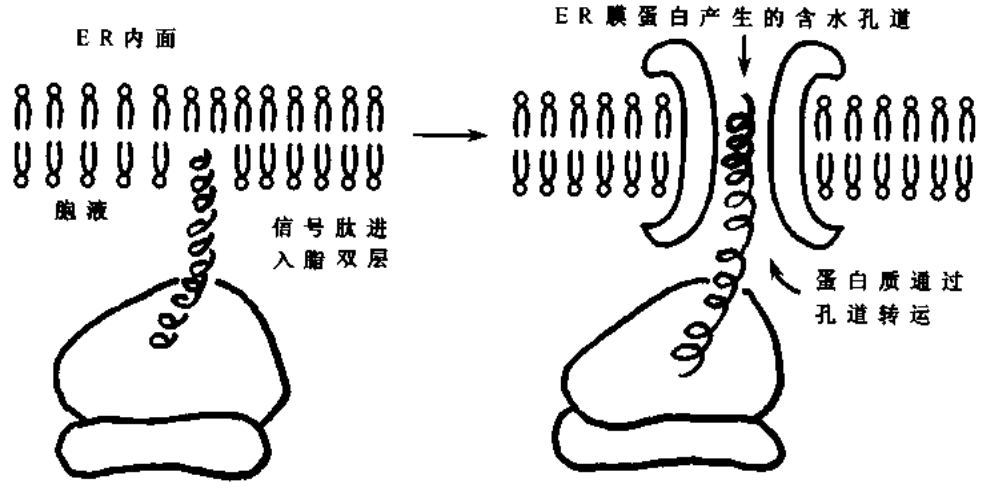


图 15－40　信号肽与脂双层的疏水环境相互作用。（仿 B. Lewin：Genes V，1994，Fig. 11－12）

中测定局部电荷的量)。当新生多肽向核糖体转运时膜上的孔道打开，由于在蛋白转运时离子不能通过孔道，所以推测转运蛋白占据了整个的孔道。但如果通过嘌呤霉素处理而释放出蛋白后的孔道就可自由渗透。如果核糖体离开膜的话，此孔道就会关闭起来，表明打开的状态需要核糖体的存在。

穿过 RE 膜的分泌蛋白和留在 RE 中的蛋白有什么区别呢？膜整合蛋白，具有疏水区，当外部结构域位于膜上的一个或另一位点时，此疏水区定位于脂双层。此疏水区提供了一个跨膜结构域(transmembrane domain)，它能使多肽在膜上漂动。它可以形成一个长 21～26aa 的螺旋跨在脂双层上(图 15－41)。

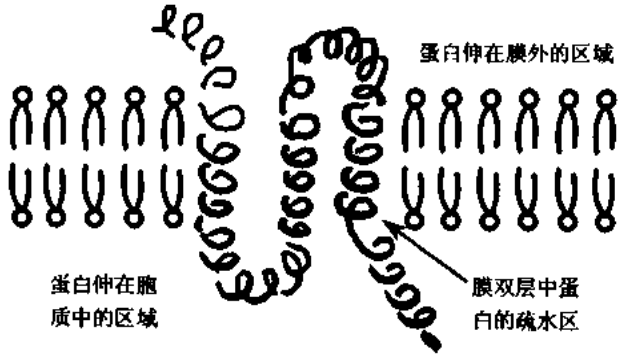


图 15－41 跨膜蛋白穿过脂双层(仿 Genes V，1994，Fig.2－5)

此信号负责蛋白插在膜中的方向。蛋白跨膜一次其 N 端在膜外侧，还在合成的 C 端位于胞液中，这种蛋白称 I 型膜蛋白，大部分膜蛋白都属 I 型。II 型蛋白的 N 端在胞液侧，C 端伸在膜外。其信号肽和锚联在一起位于 N 端。

无论 I 型还是 II 型蛋白插入膜的途径。与作为分泌蛋白的起始方法都是相同的，依赖于信号肽的共翻译功能。但留在膜中的蛋白要具有第二种停止转运(stop-transfer)信号。此信号由一组连接某些离子的疏水氨基组成。这些氨基酸作为一种锚(anchor)固着在膜上，或者使蛋白质停止转运。现在还不知其具体作用。

II 型蛋白通过信号肽开始插入膜时形成发夹环，其 N 端停在胞质这一侧。II 型蛋白的特点决定于蛋白在膜中的位置与方向。如图 15－42 信号肽进入膜，但信号肽－锚顺序使蛋白不能穿过膜。当生长的多肽形成环持续进入 ER 时，它贴在膜的脂相。

信号－锚顺序常在肽的内部，其位置决定蛋白留在胞液的部分和进入细胞器的部分。其实所有具有信号－锚顺序的 N 端顺序都是伸向胞液。通常留在胞液中的尾巴很短，约 6～30aa。当蛋白质其余部分穿越过膜时，N 端保持收缩状态。

锚肽的定义没有信号肽严格，但它适合于跨膜结构域的一般原则。锚肽的一个有趣的特点是当通过基因重组将它们加到不同的位置时，它们仍具有作为信号的功能。当位于蛋白中部的别的信号肽丢失时，锚肽也可起信号肽的作用，导致越膜。至于锚肽将肽链锚定于膜上一种可能的解释是信号肽和锚肽某些共同的越膜装置相互作用导致了蛋白转运的停止。信号肽的结合起始转运，但锚肽的存在可以停止转运。若在锚顺序的前面缺乏信号肽，那么锚肽就可以起始转运。

重组的 II 型蛋白的信号－锚肽和可切掉的导肽相似。唾液酸苷酶(图 15－43)是一个例子。像可剪切的导肽一样，氨基酸的组成比顺序更为重要。在信号－锚的末端是带有正电荷的氨基酸，中心区域是不带电荷的残基，与可剪切的导肽的疏水核心相似。若

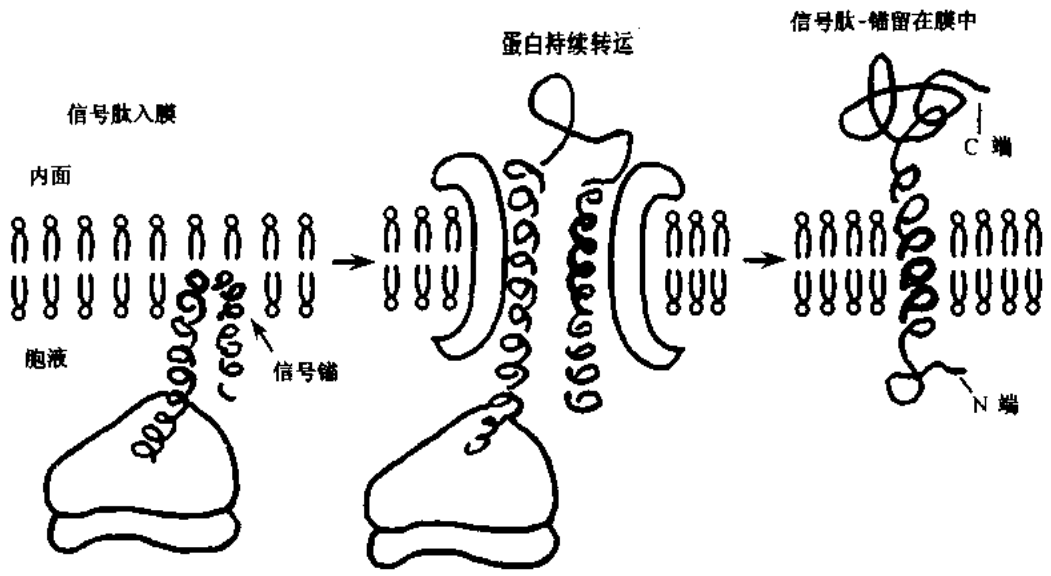


图 15 - 42　信号肽 - 锚改变蛋白质的方向，使 N 末端在胞液侧。（仿 B.Lewin:Genes V,1994,Fig.11 - 14）

经突变在核心区导入带电荷的氨基酸，就会阻止膜的插入；若突变在另一侧就会抑制锚的功能。这样蛋白就会分泌或者定位在不恰当的腔内。

锚肽周围电荷的分布对于蛋白质的方向是很重要的。较多的正电荷常发现在胞质侧（Ⅱ型蛋白的 N 端）。若通过突变除去正电荷，蛋白质的方向就会转变。电荷对方向的作用根据正电荷在内侧的原则推测，实际上锚是在胞质一侧。控制方向的是靠疏水区，它决定了蛋白的方向。

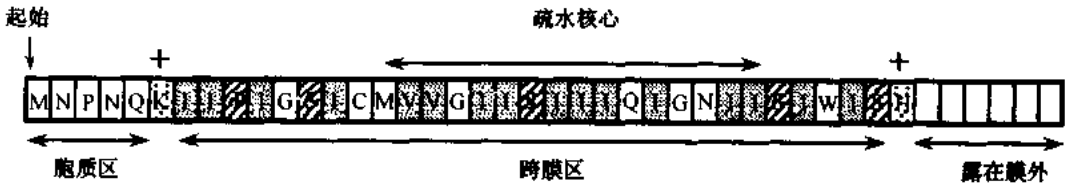


图 15 - 43　唾液酸苷酶（neuraminidase）的信号 - 锚顺序位于近 N 端，且有一个疏水核心。
（仿 B.Lewin:Genes Ⅵ,1997,Fig.10 - 24）

Ⅰ型和Ⅱ型蛋白插入膜的过程已经描述（图 15 - 40,42）。Ⅱ型蛋白有跨膜构域，而有的膜蛋白具有更为复杂的结构——**多重跨膜结构域**（multiple membrane - spanning regions）N 端和 C 端或在同侧，或在两侧，此由跨膜区是偶数还是奇数而定。

一个有趣的问题是膜整合蛋白怎样起始进入脂双层的图 15 - 40 表示蛋白通过孔道，它必须要达到孔道，由疏水跨膜区来占据它。然后这一区域必须从孔道侧面转到脂双层中，此可能要通过和孔道结合成为它的一种成分才能完成，或者通过重排使疏水区进入周围的脂双层中。

具有多重跨膜区的蛋白是怎样进入到膜中的呢？关于这个问题知道的很少，可能依赖于顺序提供的信号或锚的能力。一种模型假设：有一种可选择的系列信号肽或锚肽。转运是在第一个信号肽起始的，然后持续转运，直到第一个锚的出现。然后通过后面的信号顺序重新起始，转运直到下一个锚肽。另一种可能是：一旦一个跨膜功能域插入，后面的跨膜功能域就定位在膜上，这是它们和已插入的结构区相互作用的结果。

第五节 细菌蛋白的越膜运输

细菌中的革兰氏阴性菌如 *E.coli* 有内膜(inner membrane)，外膜(outermembrane)和周质(periplasmic space)；屏障较多，所以真正能分泌的蛋白不多，常只能分泌一些水解酶和毒素。而革兰阳性菌(如芽孢杆菌)无外膜屏障，故分泌的能力比前者强得多。

细菌蛋白是通过共翻译和翻译后机制进行越膜运输的。协同翻译常采用普遍性途径越过内膜，翻译后机制采用孔径分泌到膜外。

一、普遍性途径

细菌蛋白的分泌其机制和真核细胞的情况十分相似。转运是从细菌的细胞质越过内膜进入周质，然后(有时)再穿过外膜分泌出去。共翻译转运在 *E.coli* 中是共同性的，但不是普遍的。有的蛋白的分泌是通过共翻译和翻译后转运。

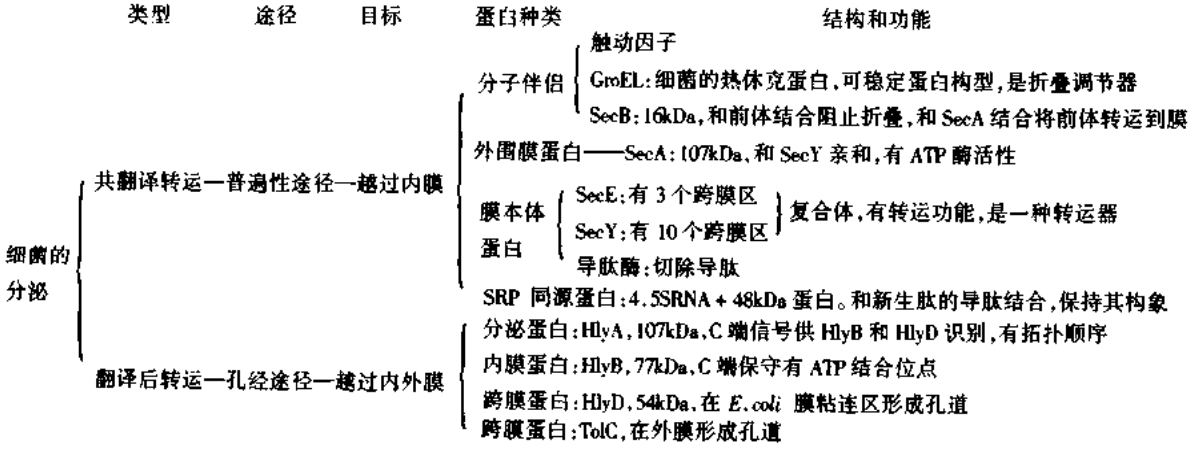


图 15-44 细菌分泌的不同途径。

细菌的输出蛋白有 N-端的导肽，它带有亲水的 N-端并连接了一个疏水核心。若 N-端的导肽发生突变就会阻止分泌。它们还会被其他基因突变所抑制，这些基因被鉴别为输出装置蛋白一些成分的基因。Sec 表示编码分泌装置成分的一系列基因。这些基因发生突变会导致很多或所有的胞外蛋白停止分泌。

如同真核的情况一样，第二个信号对于蛋白入膜后定位于恰当的位置来说是必须的。例如 *E.coli* 的 β 半乳糖苷酶的 C 端区域对于此酶越过膜进入另一侧的周质很必要。

控制折叠对于蛋白的转运也是重要的。在越膜时蛋白的构象会发生改变。例如 β-半乳糖苷酶在越膜前或在穿越细菌内膜时是以胰蛋白酶敏感形式存在的，但当它被释放到周质时期构象变成了胰蛋白酶抗性的形式。

图 15-45 表示细菌蛋白越膜转运的步骤。分子伴侣 SecB 和新生肽结合，控制其折叠。它和分泌装置 SecA 结合，而 SecA 又和膜整合蛋白 SecE/Y 结合；蛋白越过膜，蛋白 N 端可剪切的导肽被肽酶所剪切。

各种分子伴侣能够具有防止细菌蛋白成熟前折叠的效力。它们包括**触动因子**(trigger factor)(它是作为一种分子伴侣，帮助蛋白的输出)，GroEL(最初描述其作用是负责噬菌体的形态建成(morphogenesis))和 SecB(是作为 Sec 突变体的产物鉴别)。虽然

SecB 是这类蛋白中丰度最小的，但其作用是起动输出，可能它具有 2 种功能。它作为分子伴侣和新生蛋白结合。另外它与 SecA 蛋白结合。

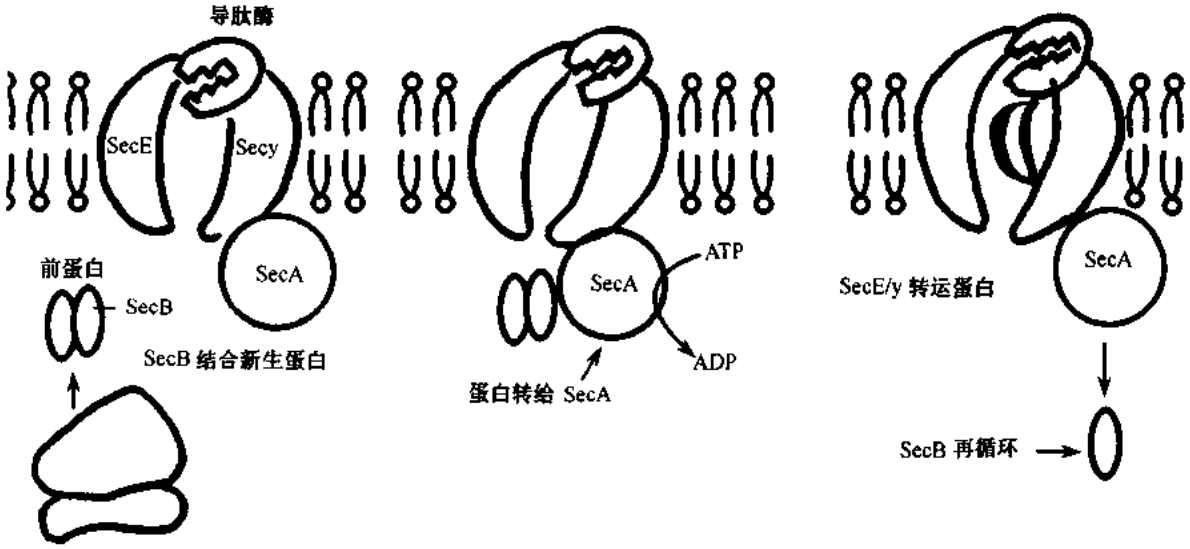


图 15 - 45　在细菌中蛋白转运需要 SecB 作为分子伴侣，它和新生蛋白结合，然后 SecB 将蛋白转给外周膜蛋白 SecA，secA 和膜整合蛋白 SecE/y 结合，转运需要水解 ATP 提供趋动力，导肽酶是一个膜整合蛋白，它可剪切导肽。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 10 - 25）

SecB 作为分子伴侣可和前体蛋白结合阻止其折叠，但它不能使已折叠蛋白的结构改变。SecB 和 SecA 的结合使它将一个前体蛋白运向膜。

SecA 是一个大的外周膜蛋白，它通过和酸性脂类及蛋白 SecY 的亲和性与膜结合。SecY 和 SecE 是膜整合蛋白（SecY 有 10 个跨膜区，SecE 有 3 个跨膜区），是多亚基复合体的一部分，提供转运的功能。

功能性的 SecA 和膜上的 SecY，SecE 等形成的复合体以高度亲和的方式相结合。这种结合激活了 SecA，使它既可以作为 SecB - 前蛋白复合体的受体并与之结合，又能作为 ATP 水解酶与 ATP 结合，并水解 ATP。虽然 SecA 的蛋白序列显示只有其 N - 端部分才具有一个较典型的 ATPase 或激酶的 ATP 结合结构域，但实际上 SecA 蛋白至少可以结合三个 ATP 分子。当 SecA 与前体蛋白一道和 SecY、E 结合时，在 ATP 的推动下，SecA 的构象发生了明显变化，促使 SecA 30kDa 的结构域插入内膜，甚至部分暴露于周质中，随之前体蛋白的约 20 氨基酸残基相应越膜，此时另一个 ATP 可能在 SecA 的第二个 ATP 结合位点水解，所产生的能量释放出前体蛋白已越膜的部分，并使 SecA 的 30kDa 结构域从膜上解离出来，与胞质中的 SecA 相互交换，以达到生理平衡，然后再进行新一轮循环，从而使前体蛋白一步一步地越过内膜。

很多的细菌蛋白是作为一种原蛋白（propotein）合成的，它的 N - 端具有可剪切的导肽。此导肽是被一种导肽酶（leader peptidase）所剪切，此酶识别各种被输出蛋白的前体，它本身是一种膜整合蛋白，位于内膜。

真核系统的突出的特点是用核糖核蛋白的 SRP 来提供信号肽的起始识别。在细菌中也有 SRP 的副本。*E. coli* 含有一个 4.5S RNA，它与核糖体结合，且和 SRP 的 7SRNA 同源。它和一个 48kDa 的蛋白结合装配。此 48kDa 的蛋白又和 SRP 的 54kDa 蛋白组分同源。此复合体在某些分泌途径中起作用，但不是所有的分泌蛋白都是如此。这些复合体的作用可能是和刚从核糖体上露出的新生蛋白的导肽相结合，使新生蛋白

保持一种恰当的构象，直到它和分泌装置的成分相互作用为止。多半这种蛋白初始时只是蛋白质合成和分泌之间的一种连接。而在真核中 SRP 又获得了翻译制动以及将新生肽导向膜的新功能。

二、孔道途径

以上是革兰氏阴性菌蛋白分泌的普遍性途径（general pathway），需要依赖于信号肽和 Sec 蛋白系统。草兰氏阴性菌还有另一种途径叫做**孔道途径**(pore pathway)如 α－溶血素(α-haemolysin，HlyA)的分泌。

Barry Holland 等(1990)已提出了 HlyA 的分泌模型(图 15－46)和 HlyA 分泌有关的基因有 4 个：*hlyA*，*hlyB*，*hlyC* 和 *hlyD*。其中 *hlyA* 是编码 HlyA 蛋白的，HlyA 分子量为 107kDa，其 N 端有较多的碱性氨基酸，而 C 端却有大量的酸性氨基酸，呈较强的极性，蛋白分子为长纤维状。其 N 端无信号肽，C 端有决定分泌的信息。在其第 230～465aa 区域内有三个疏水结构域；最后面的 > 18aa 顺序高度保守。C－端也不含信号肽，但有 27 个 aa 与分泌活性密切相关，除去后明显影响分泌。可能是与分泌有关的拓扑(topgenic)顺序。

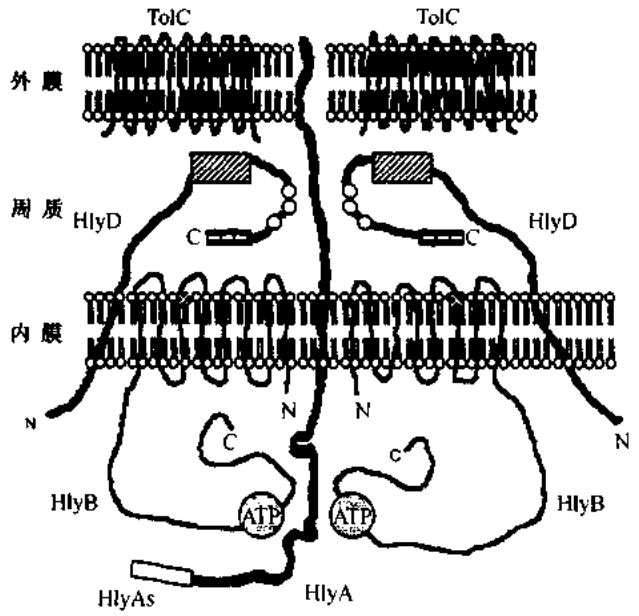


图 15－46　α－溶血素经孔道途径进行分泌的模型。(引自 Schlor，S. *et al*，Mol. Genet. 1997，256：306)

HlyB 和 HlyD 分别由 *hlyB* 和 *hlyd* 编码，二者不仅影响 HlyA 分泌而且也影响其合成。HlyB 分子量为 77kDa，位于内膜，C 端有两个高度保守区，其中一个有 ATP 结合位点，与越膜运输有关。HlyD 分子量为 54kDa，它与 HlyB 形成的复合体使 *E. coli* 膜的粘连区(membrane adhesion zones)，产生一个孔道(channel)。HlyB 与 ATP 结合，从 ATP 水解中获得能量，提供 HlyA 分泌的需能量。HlyA 的 C 端可识别 HlyB 和 D 复合体。

M13 的外壳蛋白(coat protein)合成后可插入膜中，这也是翻译后转运的另一个例子。

以上表明导肽是分泌中（普遍性途径）不可缺少的，但并不是惟一决定因素，在非分泌蛋白上加上信号肽后，常常并不引起分泌，因此在构建分泌蛋白表达载体时，还应考虑到其他因素(图 15－46)。

第六节　翻译后的加工

前体蛋白是没有活性的，常常要进行一个系列的翻译后加工，才能成为具有功能的成熟蛋白。加工的类型是多种多样的，一般分为四种：N 端 fMet 或 Met 的切除，二硫键的形成；化学修饰和剪切。

一、N 端 fMet 或 Met 的切除

原核生物的肽链，其 N 端不保留 fMet，大约半数蛋白由**脱甲酰酶**(deformylase)除去甲酰基，留下 Met 作为第一个氨基酸；在原核及真核细胞中 fMet 或者 Met 一般都要被除去，这是由**氨肽酶**(amino peptidase)水解来完成的。水解的过程有时发生在肽链合成的过程中，有时在肽链从核糖体上释放以后。至于脱甲酰还是除去 fMet 常与邻接的氨基酸有关。如第二氨基酸是 Arg，Asn，Asp，Glu，Ile 或 Lys 以脱甲酰基为主，如邻接的氨基酸是 Aly，Gly，Pro，Thr 或 Val 则常除去 fMet。

二、二硫键的形成

两个半胱氨酸相距较远，硫氢基可以氧化成二硫键，产生 mRNA 中没有相应密码子的胱氨酸。

三、化学修饰

化学修饰是蛋白质加工的重头戏，修饰的类型很多，包括**磷酸化**(如核糖体蛋白的 Ser，Tyr 和 Trp 残基常被磷酸化)；**糖基化**(如各种糖蛋白)；**甲基化**(如组蛋白，肌蛋白)，**乙基化**(如组蛋白)，**羟基化**(如胶原蛋白)。还有各种辅基如大卟啉环结合在叶绿蛋白和血红蛋白上。

糖基化是真核细胞中特有的加工，这些蛋白常和细胞信号的识别有关，如受体蛋白等。

1. 折叠

在 ER 腔中折叠和修饰是有关，糖的连接对于正确的折叠是十分必要的。蛋白二硫异构酶 (protein disulfide isomerase，PDI) 可以改变二硫键，影响到折叠，它和特殊的 ER 蛋白的结合是必须的，此酶的某些活性或全部的活性可能是作为 ER 中的一种复合体的形式来实现的。即在越膜位点和蛋白结合才能发挥它的功能。通过对折叠和寡聚物产物的计算表明折叠需要一种酶来催化，使其在细胞中迅速发生。

另一个与折叠功能有关的蛋白是 BiP，它是分子伴侣 Hsp70 家族的一个成员。BiP 促使寡聚物的形成，帮助 ER 腔中蛋白的折叠。ER 可能含有各种辅助蛋白，它们的功能是识别蛋白折叠的形态，以帮助这些蛋白产生一种构象，使其可迅速地转运到下一个目标。

大部分膜上的糖蛋白，都是寡聚物，一般在 ER 中寡聚，然后迅速地从 ER 转到高尔基体，但不装配亚基或者防止蛋白错装配。错折叠(misfolding)的蛋白常和 BiP 相联，在这

种情况下他们会被降解掉，若折叠正确，那么就可以转运到高尔基体或继续转运。

BiP 有两个功能：①帮助转运蛋白折叠；②切除错折叠蛋白。据推测 BiP 识别一定的肽顺序，而在成熟蛋白正确折叠的构象中这种顺序不在外围，而刚进入 ER 腔中的新生肽这些顺序露在外面，就会附着到 BiP 上。当某种蛋白发生了错折叠或变性的话，这种顺序就露在外表面上，所以可被 BiP 识别和结合。但现在还不清楚一个持久的错折叠的蛋白，怎样作为降解的目标，它们是否和 BiP 连接。

2. 在 ER 中的糖基化及修整

实际上所有的分泌蛋白和膜蛋白几乎都是被糖基化的蛋白。糖基化有两种类型：①糖蛋白是由寡糖连接在 Asp 的氨基的形成的（图 15－47），连接的链叫 N－糖苷键。②寡糖连接在 Ser、Thr 或羟基－Lys 的羟基上（O－糖苷键）。N－糖苷键是在 ER 开始，而在高尔基体中进一步完成；O－糖苷键（O－linked glycosylation）的形成仅发生在高尔基体中。

N－糖基化可分为 3 步，各种 N－连接的寡糖都是在 ER 中开始加上的（图 15－48），途径也相同。一个寡糖含有 2 个 N－乙酰－糖胺（N－acetyl－glucosamine），9 个甘露糖和 3 个葡萄糖，在一种特异的脂—多萜醇（dolichol）上形成，多萜醇磷酸酯即是携

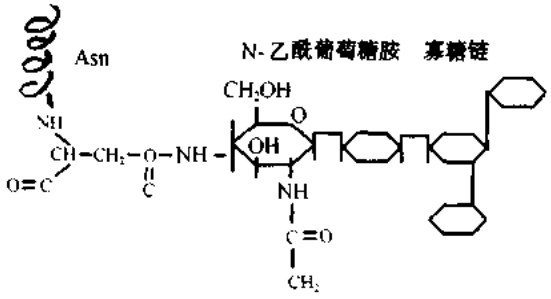


图 15－47 N－连接的糖蛋白有一个寡糖链连接在 Asn 的氨基上。
（仿 B. lewin: Genes Ⅵ, 1994, Fig. 1－10）

带糖的载体。多萜醇是一种高度疏水的酯，位于 ER 膜中，其活性基团面向着 ER 腔，寡糖是由单个的糖连接而构成，它通过焦磷酸和多萜醇连接。寡糖作为一个单位在与膜结合的糖基转移酶（goycosyl transferase enzyme）的作用下从多萜醇上转移到靶蛋白上。酶的活性位点也是露在 ER 腔中。受体基团是位于 Asn－X－Ser 或 Asn－X－Thr（X 是除 Pro 以外的任何氨基酸）中的 Asn，当新生肽进入 ER 时，它一旦被识别后立即作为靶顺序暴露在腔中。有些寡糖的修整是在 ER 中进行，修整后再进入高尔基体。寡糖结构完成分为两类，一类在转运到 ER 时，另一类是在越膜转运到高尔基体时。究竟属于何种类型这要取决于甘露糖。甘露糖只是在 ER 中加上，随后可能还要被修整。在 ER 中被切去的单糖的数目是不同的。ER 中的甘露糖苷酶（ER mannosidase）很快地附着到第一个甘露糖上，附着下三个较慢。

高甘露糖寡聚糖（high mannose oligosaccharides）含有的残基是在 ER 中加上的，修整的过程如图 15－49 所示。寡糖加上后几乎立即又从蛋白上被切掉。3 个葡萄糖残基被 ER 中的葡萄糖苷酶（glucosidases Ⅰ和Ⅱ）切除掉，ER 中的甘露糖苷酶再从蛋白上切下 2～4 个甘露糖，在 ER 中产生 3 寡糖的最终结构。

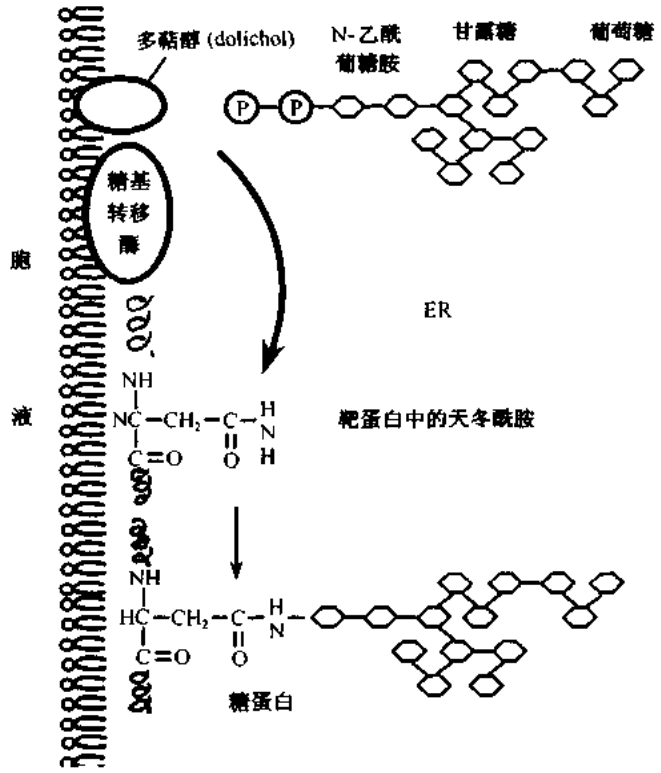


图 15－48　多萜醇上的寡糖通过糖基转移酶转移到靶蛋白的 Asn 上。
(仿 B. Lewin: Genes V, 1994, Fig. 11－17)

3. 在高尔基体中的进一步加工

复合寡聚糖 (complex oligosacchrarides) 是在高尔基体中进一步修整和加上的糖的残基。顺序如图 15－50 所示。第一步是通过高尔基体的甘露糖苷酶Ⅰ修整甘露糖残基。然后单个的糖基由 N－乙酰－葡萄糖胺转移加上，按着由高尔基体甘露糖苷酶Ⅱ继续切除甘露糖残基。在这一位置上寡糖通过**内－糖苷酶 H**(endo-glycosidase H, Endo H)使寡糖对降解产生抗性。对 Endo H 的易感性(susceptiblity)被用来作为一种测验，从而测定一个糖蛋白是处于什么阶段。

在高尔基体的修饰中都会产生**内部核心** (inner core)，它是由 NAc－Glc·NAc－Glc.Man3 构成，最后要被切掉。末端区域(terminal region)加在内部核心下。末端区域的残基包括 NAc－Glc, Gal 和唾液酸(N－乙酰－神经氨[糖]酸)。此加工的路径和糖基化是高度有序的，而且有两种类型的反应。一种糖残基的加入对于另一种糖基的剪切可能是必要的。如在最终剪除甘露糖之前要加上 NAc－Glc。

现在还不知道加工过程中各种蛋白的降解，加工的模式及糖基化的信号是什么。据推测此信号在肽链的结构中，而不可能在寡糖中，因为 N－糖苷键开始形成都是加上相同的寡糖。

四、剪切

很多的前体蛋白要经过剪切后方可成为成熟的蛋白。

在原核生物中常常产生一种多蛋白的前体要经剪切后才能成为成熟的蛋白，如反

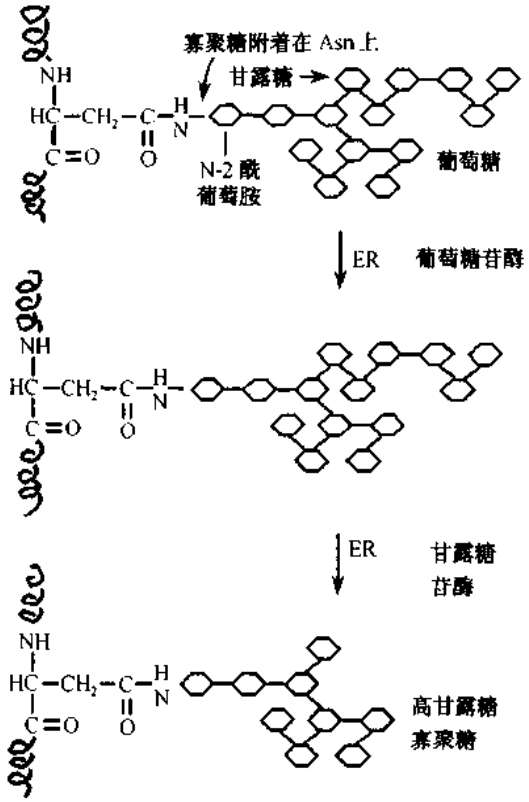


图 15－49 糖在 ER 中以一定的次序被切除。(仿 B. Lewin: Genes V, 1994, Fig. 11－18)

转录病毒中有 3 个基因 *gag*, *pol* 和 *env*, 其中 *pol* 基因长约 2900nt, 其产物经剪切后产生反转录酶, 内切酶和蛋白酶三种蛋白。其他两个基因的产物也要经过加工才能产生核心蛋白和外壳蛋白。

在真核生物中有些蛋白要经过剪切才能成为有活性的成熟蛋白, 最有名的例子是高等生物的胰岛素, 它是一种分泌蛋白, 具有信号肽。新合成的**前胰岛素原**(preproinsulin), 在 ER 中切除信号肽变成了胰岛素原(proinsulin), 它是单链的多肽, 由 3 个二硫键将主键连在一起, 弯曲成复杂的环形结构。分子由 A 链(21aa)B 链(31aa)和 C 链(33aa)三个连续的片段构成。当转运到胰岛细胞的囊胞中, C 链被切除, 成为由 A, B 两条分开的链由 3 个二硫键连结成成熟的胰岛素(图 15－51)。

蛋白内含子又称为**内蛋白子**(intein)是近年发现的一种新的翻译后加工的产物。它是 1994 年由 Perler 等首先提出的。蛋白外显子又称为**外蛋白子**(estein)。内蛋白子的基因不是单独的开放阅读框 (ORF) 它是插入在外蛋白子的基因中和内含子的区别在于它可以和外蛋白子的基因一道表达, 而不是 mRNA 阶段被切除, 产生前体蛋白以后再从前体中被切除掉, 余下的外蛋白子连接在一起成为成熟的蛋白 (图 15－52), 这也不同于胰岛素的情况, 胰岛素的 A 键和 B 键本身是不相连接的, 仅仅通过二硫键将两个片段连在一起。

内蛋白子的基因除了上下代的垂直传递以外, 还可以通过基因的转变(gene coversion)而进行水平传递, 被称之为**内蛋白子归巢**(intein homing)。

现发现的内蛋白子有 7 种, 多分布于酵母和微生物中, 内蛋白子的分子量为 40～60kDa, 有高度保守的末端氨基酸: N－端常为 Cys 或 Ser, C－端总是 His－Asn。内蛋白子未剪切前称融合内蛋白子(fused intein), 可催化前体的自我剪接反应; 剪切后的内蛋

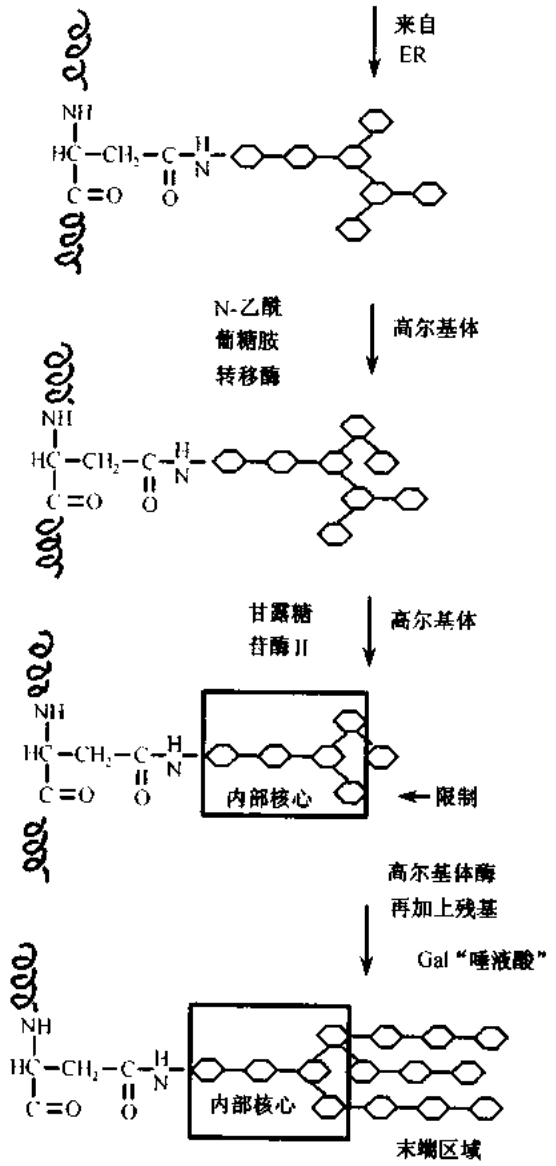


图 15－50　糖蛋白在高尔基体中的进一步加工。(仿 B. Lewin: Genes Ⅴ, 1994, Fig. 11－19)

白子称**游离内蛋白子**(free intein)，可作为**归巢内切酶**(homing endonuclease)参与内蛋白子的归巢。它可识别 dDNA 上内蛋白子基因在外蛋白子基因子的插入位点。

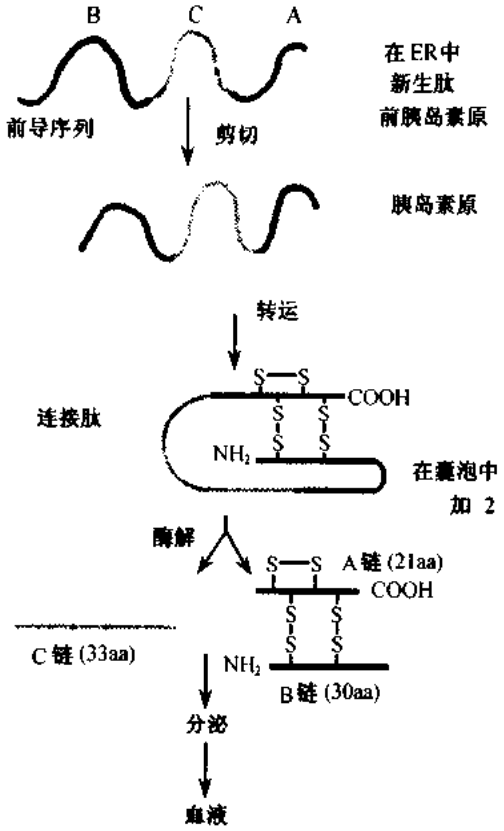


图 15－51　胰岛素的翻译后加工。

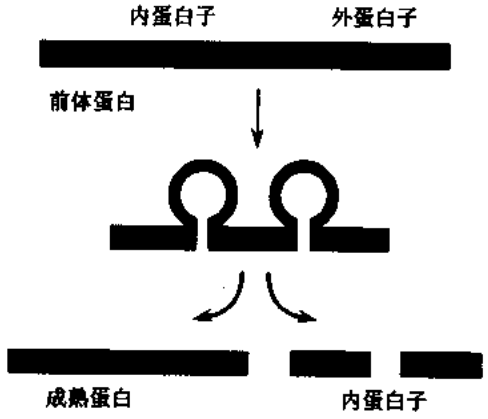


图 15－52　内蛋白子的剪切。

第十六章

原核生物基因表达的调控

基本概念

原核生物基因表达的调控可以在DNA、转录和翻译三个不同层次进行，但转录水平的调控是最主要的方式，也是最经济、最有效的方式。

操纵子是原核转录调控的主要形式，相关的基因排列成簇，由一个调节蛋白所控制，一开俱开，一闭俱闭，对环境条件的改变作出相应的反应。

操纵子主要可分为三大类型，一种是调节分解代谢的操纵子，它们都属于诱导型，并同时受到cAMP－CAP的调节。分解代谢的底物常为小分子诱导物，不具衰减作用。另一种是调节合成代谢的操纵子，它们都属于阻遏型，不受cAMP－CAP的影响。第三种类型具有衰减子（如 *trp*，*his*，*phe*，*leu*，*thr*，和 *ilv*），最终产物为辅阻遏物。

噬菌体的时序调控一般都是在转录水平控制的。不同的噬菌体采用的策略不同，常见的有三类：①σ因子的替换；②核心酶的修饰；③核心酶的替换。其中σ因子的替换更为普遍。

衰减作用要具备三个重要的条件，①前导序列中要有相应氨基酸的密码子；②要具有4组配对区；③转录和翻译必须偶联。真核生物的转录和翻译是分别在核和胞质中进行，不能偶联，所以也不存在这种形式的调节。

DNA水平的调节主要通过DNA重排，例如在原核中沙门氏菌的相变位和Mu噬菌体G片段的倒位都是通过DNA的重排来改变转录产物的序列，以此来调节宿主范围。

翻译水平的调控也可分为三种不同阶段（起始，延伸和终止）进行调控。

翻译水平的调控较为重要的是反义RNA的调控，二级结构对翻译的调控，蛋白质的自主调控和严紧反应。

人们根据反义RNA的原理已建立了反义技术，应用于遗传病、肿瘤、病毒病、寄生虫病的防止研究以及抗性植物、特殊花卉、迟熟果蔬的培育。

生物并不是在一个恒定的环境条件下生活，繁殖。通过进化的过程，生物建立了一种适应于环境改变的能力，其方法之一就是改变其基因的活性，这样基因产物的合成对新的条件正好适合，生物在这样的环境中能正常的的生长和繁殖。已经在自由生活的细菌中证实了这点。

在一个原核细胞中一系列基因产物合成的改变要涉及到调控机制，即基因表达的控制，也就是在基因的拷贝数，转录，翻译等不同层次来调节蛋白质的合成。在每一层次又以不同的形式来进行调节。一般情况下，根据细胞或生物的需要，其活性可以受到**调节型基因**（regulated genes）的控制。一个有机体还有大量的另一些基因，这些基因的产物维持了生长和细胞分裂的正常功能，不论环境条件如何，在生长的细胞中这些基因总是处于活性状态，这些基因就叫做**组成型基因**（constitutive genes）也叫**看家基因**（Housekeeping genes），例如编码合成蛋白的基因，合成蛋白时需要的一系列酶的基因及葡萄糖代谢所需的基因。探索各种基因在什么水平上进行调控是重要的。若环境突然变得对正常细胞功能不利的话，那么就会通过特殊调节机制来降低表达，这样调节型基因和组成型基因间的差别也就淡化了。

在这一章中将系统介绍原核生物特别是在细菌和噬菌体系统中基因表达的调节。在不稳定环境中生活的野生型细菌，食物的供给并不能得到保证或得不到一定类

型的营养。细菌要能活下来必须调节细胞活性来适应改变了的条件。在选择的压力下,细菌产生了各种基因打开和关闭机制。

由于噬菌体是寄生在细菌中的,它又依赖于宿主菌产生的活性。所以它们必须以一种谨慎的方式来完成自己特有的繁殖周期。本章将要介绍噬菌体的基因调节及噬菌体 λ 殖解周期和溶源周期的选择调节。

原核基因表达的调控可以在转录和翻译水平上以不同形式进行调节,转录水平的调节是最有效,最经济的方式,也是最主要的调节方式。

第一节 转录水平的调控

一、操纵子模型

1961 年法国科学家 Jacob 和 Monod 提出了一个控制细胞基因表达的模型称为**操纵子**(operon),此模型的提出使基因概念又向前迈出了一大步。人们已认识到基因的功能不是固定不变的,而是可以根据环境的变化进行调节。随之人们发现无论是真核还是原核生物转录调节都涉及到编码蛋白的基因和 DNA 上的元件。这一发现获得了一九六五年诺贝尔奖。

一个基因就是一段编码有功能产物的 DNA 序列。基因的产物可以是蛋白质或是 RNA(如 tRNA 和 rRNA)。基因的重要特点是在某些情况下其产物能从合成位点扩散开去作用别的位点。

DNA 元件是 DNA 上一段序列,它不能转变成任何其他的形式,但它作为一种**原位**(*in situ*)序列具有特殊的功能。由于它只能作用同一条 DNA,因此称**顺式作用元件**(*cis*-acting element)。

基因可以根据它们的产物分成不同的类型。编码细胞必要蛋白,如酶或结构蛋白的基因称为**结构基因**(structural genes)。这类基因在细胞中占绝大部分,承担着细胞各种蛋白的结构和功能。编码调节蛋白的基因称**调节基因**(regulator genes)。调节蛋白可调节其他基因的表达。由于调节基因的产物可以自由地结合到其相应的靶位点上,因此被为反式作用(*trans*-acting)**因子**。

调节的关键是调节基因编码调节蛋白,此蛋白通过和 DNA 上特殊位点的结合来控制转录。顺式作用元件被反式作用因子识别可以以"正"的或"负"的形式调节靶基因,顺式作用位点通常总是在靶基因的上游。

1. 调节基因和结构基因

当在细菌培养基中加入某些底物(如 Lac)时可以诱导相应基因的表达。这些基因称为**可诱导的**。诱导基因的底物称为**诱导物**(inducer),它是一类小分子**效应物**(effectors),与很多被调节基因的表达控制有关。

可诱导的基因和**组成型基因**两者有所不同。对于可诱导的基因来说它只在当特殊的分子和基因附近特殊位点发生作用时才能转录。这种 DNA 上的特殊位点称为**控制位点**(controlling site)。这种调控涉及到诱导物和调节蛋白。在特殊位点受到调节的作用下,基因通过诱导被打开,产生 RNA,并合成相应的酶。基因仅对诱导物反应

产生基因产物的现象叫做**诱导**(induction)。

组成型基因在细胞生长和分裂时总是持续地表达。它们的启动子也总是可以和RNA聚合酶结合，所以基因的转录是持续的，并恒定供应必要的基因产物。这些组成型基因是在启动子的控制之下。组成型基因可能也有控制位点，但这些位点只能控制合成的速率。在一定的条件下，组成型基因的表达也需要减少或阻断。这些条件包括热休克和饥饿状态，即培养基中氨基酸耗尽或在缺少某种氨基酸的培养基中生长。

2. 正调控和负调控

正、负调控系统是根据调节蛋白缺乏时，对操纵子的影响来命名的(图 16－1)。基因在**负控制**的情况下可被阻遏蛋白关闭掉，当阻遏蛋白缺乏或失活时基因才能打开。任何一种干扰基因表达的作用都属于负控制，其机制或是阻遏蛋白和DNA上的特异位点是相结合，阻止DNA聚合酶起始转录，或是阻遏蛋白和RNA结合阻止翻译的起始。

负调控提供了一个保险机制，如调节蛋白失活的话，这个系统仍有功能，因而细胞中仍合成相关的一系列酶。我们不难设想这可能是一种进化的结果。原始的系统其表达可能是组成型的，以后细胞产生一种干预这种表达的机制，获得了一种选择优势，特异地控制表达以提高效率，减少浪费，演化为负调控系统。

基因在**正调控**的情况下只有在调节蛋白处于活化状态时才能表达。对于一个正控制操纵子来说其机制正好和负控制相反，调节蛋白的作用不是阻止起始，而是帮助起始。它和DNA以及RNA聚合酶相互作用来帮助起始。正调节的蛋白对小分子作出反应，一般称为**激活物**(activator)。另一类正调控(如 σ 因子)可以改变启动子的选择，抗终止因子改变对**终止子**(terminators)的选择。

根据操纵子对调节基因表达的小分子所作出反应的特点，分为**可诱导**(inducible)和**可阻遏**(repressible)两类。

例如在细菌的培养基中只有在加入代谢的诱导物之后，细菌才产生一种酶。这就是可诱导的操纵子；如在培养基中加入足量的某些合成物的成分，就可以阻遏合成时所需的一系列酶的产生，这就是可阻遏操纵子。可诱导操纵子只有在小分子诱导物存在的情况下才有功能；可阻遏的操纵子只有在小分子**辅阻遏物**(corepressor)缺乏的情况下才有功能。

用于可阻遏系的术语是：操纵子的活性状态称为去阻遏(derepressed)，与"**被诱导**"的意思相同。若操纵子发生突变而不能去阻遏的活有时被称为**超阻碍**(super-repressed)，也就是"**不可诱导**"的意思。

无论是正控制还是负控制都是通过利用调节蛋白与小分子诱导物或辅阻遏物之间的相互作用来完成诱导和阻遏。图 16－1 表示 4 种相似的控制类型，每一个诱导物使阻遏物失活或活化激活蛋白时产生诱导；当辅阻遏物活化阻碍蛋白或使激活蛋白失活时产生阻遏。

失活的调节蛋白的遗传学效应用于区分正、负控制系统。调节蛋白失活产生负调控系统的隐性组成型(或去阻遏)表达，在另一方面，使激活蛋白失活的突变导致正控制系统产生隐性的不可诱导性或超阻遏。

色氨酸操纵子是可阻遏系统的例子之一，Trp是由一系列生物合成酶类催化的生化反应的终产物。Trp酶的活化及合成是由细胞中Trp的含量来控制。

经典的终产物**反馈抑制**适用这些酶类：色氨酸合成途径的第一个酶的催化活性被

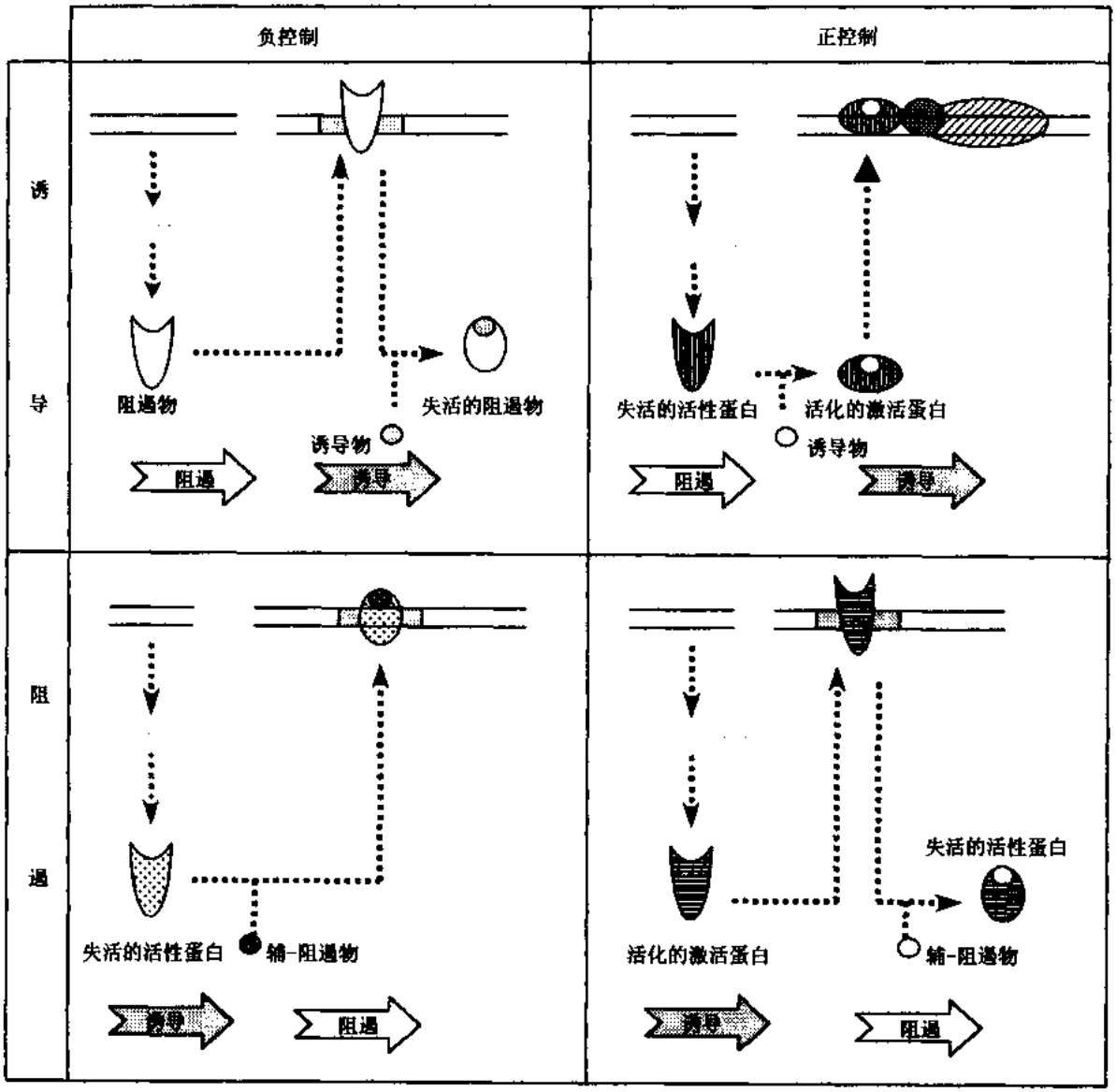


图 16-1 原核生物结构基因的 4 种表达调控类型。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 12-21)

终产物 Trp 所抑制,也就是说当细胞中含有足够的 Trp 时,它就能通过抑制合成途径的起始来切断此氨基酸分子的进一步合成。

Trp 的功能就是作为一种辅阻遏物来活化阻遏蛋白。这是典型的阻遏机制的例子(图 16-1),当 Trp 供应充足时,由于阻遏蛋白-辅阻遏物复合体与**操纵基因**(operator)结合,操纵子被阻遏。当 Trp 供应不足时,辅阻遏物失活,结合在操纵基因上的复合体减少,操纵子重新打开。

阻遏蛋白的去除导致 Trp 启动子上的起始频率增加了约 70 倍,甚至在阻遏的条件下,此结构基因可持续地**本底**(basal)**表达**,或以**被阻遏水平**(repressed level)**表达**。在这种操纵基因上的阻遏效率要比 *lac* 操纵子低得多(*lac* 操纵子本底表达的水平只有诱导状态的 1/1000)。

小分子可以改变调节蛋白的构象而产生诱导和阻遏的现象。一个例子是 OxyR,通过过氧化氢诱导基因的转录激活物。Oxy 蛋白可直接被氧化作用所活化,这样它就为

氧化应激反应提供了一个敏感源。另一个类似的例子是调节蛋白的磷酸化。

二、乳糖操纵子

1. 结构和功能

细菌相关功能的结构基因常连在一起，形成一个基因簇。它们编码同一个代谢途径中的不同的酶。一个基因簇被共同调控，一开俱开，一闭俱闭。也就是说它们形成了一个被调控的单位。其他的相关功能的基因也包括在这个调控单位中，例如编码透过酶的基因，虽它的产物不直接参与催化代谢，但可以使小分子底物转运到细胞中。

乳糖分解代谢相关的三个基因，*lacZ*、*Y*、*A* 就组成典型的基因簇。它们的产物可催化乳糖的分解，产生葡萄糖和半乳糖。图 16－2 表明这个基因簇的组成。它们具有顺式作用调节元件和反式作用调节基因。三个结构基因的功能是：

lacZ 编码 β－半乳糖苷酶，此酶由 500kDa 的四聚体构成，它可以切断乳糖的半乳糖苷键，而产生半乳糖和葡萄糖(图 16－3)。

lacY 编码 β－半乳糖苷透性酶，这种酶是一种分子量为 30kDa 的膜结合蛋白，它构成转运系统，将半乳糖苷运入到细胞中。

lacA 编码 β－半乳糖苷乙酰转移酶，其功能只将乙酰－辅酶 A 上的乙酰基转移到

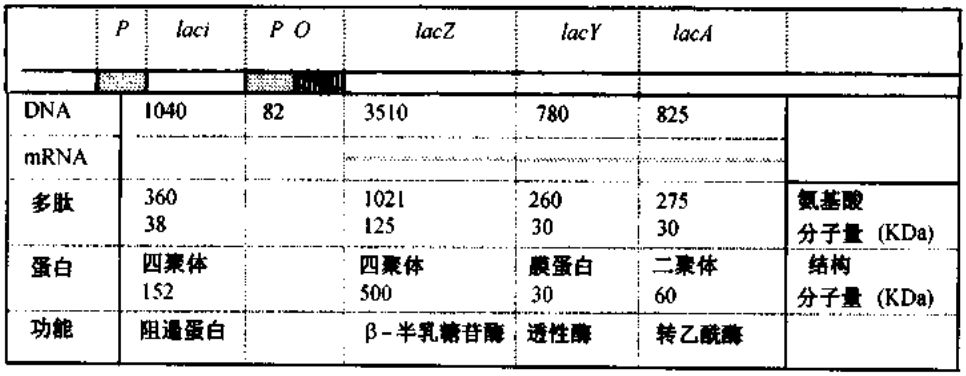


图 16－2　乳糖操纵子的结构和功能。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－3)

β－半乳糖苷上。

无论是 *lacZ* 发生突变还是 *lacY* 发生突变都可以产生 *lac*⁻型表型，这种 *lac*⁻表型的细胞不能利用乳糖。*lacZ*⁻突变体中半乳糖苷酶失去活性，直接阻止了乳糖的代谢。*lacY*⁻突变体不能从膜上吸取乳糖。

完整的调节系统包括结构基因和控制这些基因表达的元件，形成了一个共同的调节单位，这种调节单位就称为操纵子(opron)。操纵子的活性是由调节基因控制的，调节基因的产物可以和操纵子上的顺式作用控制元件相互作用。

lacZ、*Y*、*A* 基因的转录是由 *lacI* 基因指令合成的阻遏蛋白所控制。*lacI* 一般和结构基因相毗连，但它本身具有自己的启动子和终止子，成为独立的转录单位。由于 *lacI*

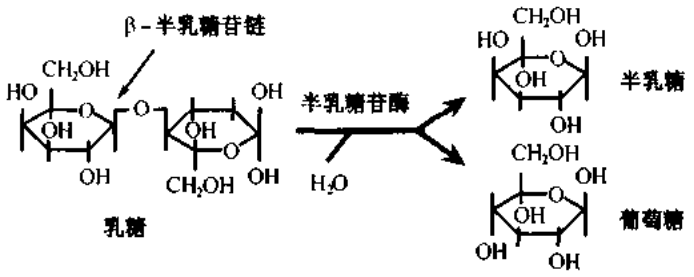


图 16-3 乳糖水解成半乳糖和葡萄糖。

的产物是可溶性蛋白，按理说无需位于结构基因的附近，而能够分散到各处或结合到分散的 DNA 位点上(这是典型的反式作用调节物。)

通过突变的效应可以将结构基因和调节基因相区别的，结构基因发生突变，细胞中就失去这些基因合成的蛋白。但是调节基因发生突变会影响到它所控制的所有结构基因的表达。调节蛋白的突变的结果可以显示调节的类型。

lac 基因簇受到负调节(negative regulation)。它们的转录可被调节蛋白所关闭。若调节蛋白因突变而失活就会导致结构基因组成型表达。表明调节蛋白的功能是阻止结构基因的表达，因此称这些蛋白为"阻遏"蛋白。

乳糖操纵子的阻遏蛋白是由 4 个亚基(38kDa)组成的四聚体。一个野生型细胞中大约有 10 个四聚体。调节基因转录成单顺反子的 mRNA，它和操纵子的比率与 RNA 聚合酶和启动子之比是相似的。

lacI 的产物称为 *lac* **阻遏物**(*lac* repressor)，其功能是和 *lacZ*、*Y*、*A* 基因簇 5′端的**操纵基因**(O_{lac})结合，操纵基因位于**启动子**(P_{lac})和**结构基因**(*lac2yA*)之间。当阻遏物结合在操纵基因上时就阻碍了启动子上的转录起始。图 16-4 表明 *lac* 结构基因开始的区域。O_{lac} 从 mRNA 转录起始点的上游 -5 处延伸到转录单位 +21 处。这样它和启动子的末端发生重叠。新近的观点认为阻遏物影响了 RNA 聚合酶的功能，从操纵基因和启动子二者相关位置来看阻遏物结合在 DNA 上会阻碍 RNA 聚合酶转录结构基因。但我们必须注意其他一些操纵子上的操纵基因其位置和乳糖操纵子并不相同，因而阻遏蛋白可以通过多种方式阻断转录操纵基因转录。

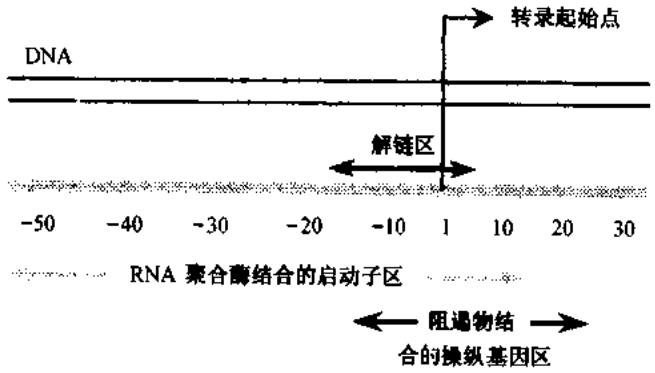


图 16-4 在 *lac* 操纵子上阻遏物结合区和 RNA 聚合酶结合区在转录起始点附近重叠。
仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12-4)

2. 阻遏蛋白的活性受到小分子诱导的控制

细菌对环境的改变必需作出迅速的反应。营养供给随时都可能发生变化。要能得以幸存必须具有可以变换不同代谢底物的能力。单细胞真核生物也同样生活在不断变化环境中；而更为复杂的多细胞生物都具有一套恒定的代谢途径，受外部环境影响较小。

细菌需要灵活性，也需要经济，因为遇到合适的环境就大量消耗营养对其本身也是不利的。在缺乏底物时就不必要合成大量相关的酶类，因此细菌产生了一种调节机制，即在缺乏底物时就阻断酶的合成途径，但同时又作好了准备，一旦有底物存在就立即合成这些酶。

特殊底物的存在导致了酶的合成，此现象称为**诱导**(induction)。这种类型的调控广泛存在于细菌中，在较低等的真核生物(如酶母)也有这种情况。*E. coli* 的**乳糖操纵子**提供了这种调控机制的典型范例。

当 *E. coli* 生长在缺乏 β－半乳糖苷的条件下是不需要 β－半乳糖苷酶的，因此细胞中含量很低，大约每个细胞不高于 5 个分子，当加入底物后细菌中十分迅速地合成了这种酶，仅在 2～3 分钟之内酶就可以产生并很快增长到 5000 个分子/每个细胞。酶的浓度将达到细胞总蛋白的 5－10%。如在培养基中除去底物，那么酶的合成也就迅速停止，恢复到原来的状态。

图 16－5 表明如果原来培养基中无乳糖，也无葡萄糖，那么细胞只在很低的基本水平合成 β－半乳苷酶和透性酶。当加入 Lac 后，*E. coli* 的 *lac*$^+$ 细胞很快大量合成以上两种酶。进一步用 ^{32}P 标记 mRNA 作杂交实验(用 λ*lac* 中的取得的 DNA，与加入乳糖后不同时间内产生的 ^{32}P-mRNA 进行分子杂交)，结果表明加入的乳糖能激发 *lac* 的 mRNA 的合成。*lac* mRNA 极不稳定，其半衰期仅有 3 分钟，当诱导物一除去转录立即停止，在很短的时间内所有的 *lac* mRNA 即被降解掉，细胞内的含量恢复到基础水平。

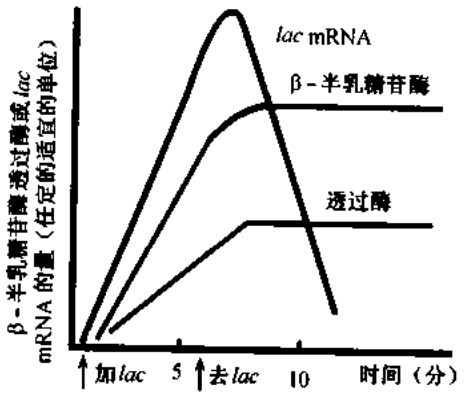


图 16－5　lac 诱导的结果。(Davis Frefelder: Molecular Biology, 1983, Fig. 14－1)

β－半乳糖苷酶和透性酶合成是和 *lac* mRNA 同时被诱导的，但当除去诱导物时在细胞中 β－半乳糖苷酶和透性酶要比 *lac* mRNA 稳定，因此酶的活性在一段较长的时间内保持被诱导水平。这种对营养供给发生改变作出迅速反应的调控类型，提供了代谢新底物的能力。比如 *E. coli* 的 Trp 的合成是通过 Trp 合成酶的作用。如果在细菌生长的培养基中加入 Trp 的话，那么立即停止 Trp 合成酶的生产。这种作用称为**阻遏**(repression)效应。它使细菌避免合成多余的物质。

在细菌中同时存在着诱导和阻遏的现象。诱导调节细菌分解底物供给生长的能力。阻遏调节细菌合成代谢产物的能力。无论是酶作用的小分子底物的调节，还是酶

活性的产生，它们的启动是独自的。小分子底物称为**诱导物**(inducers)，某些物质能阻止酶合成它们本身，此物质就称**辅阻遏物**(corepressors)。

诱导和酶阻遏是高度特异的，只有底物/产物或紧密相关的分子才能起作用，但小分子的活性并不依赖于和靶酶的相互作用。某些诱导物与自然的β－半乳糖苷酶相似，但并不能被酶分解，比如异丙基－β－D－硫代半乳糖苷(isopropylthiogalactoside，**IPTG**)。其半乳糖苷键中用硫代替了氧，失去了水解活性，但硫代半乳糖苷和同源的氧代化合物与酶位点的亲和力相同(图 16－6)，**IPTG** 虽不为β－半乳糖苷酶所识别，但它是 *lac* 基因簇十分有效的诱导物。

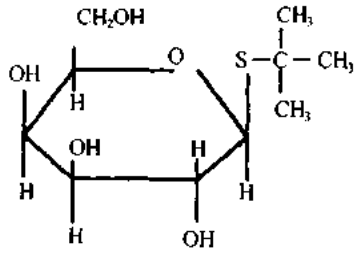


图 16－6　异丙基－β－硫代半乳糖苷的分子结构。

能诱导酶的合成，但又不被分解的分子，称为**安慰诱导物**(gratuitous inducer)。由于乳糖虽可诱导酶的合成，但又随之分解，产生很多复杂的动力学问题，因此人们常用安慰诱导物来进行各种实验。它的存在表明一个重要的问题，就是这个控制系统必须具有某种成分，它不同于靶酶，但能识别合适的底物；而它的这种识别相关底物的能力也不同于酶。

对诱导物作出反应的这种成分就是阻遏蛋白，它由 *lacI* 编码，其作用是控制 *lacIYA* 结构基同的转录，对环境作出反应。三个结构基因转录成单个的**多顺反子 mRNA**。阻遏蛋白的活性状态决定了此启动子是打开或关闭。在缺乏诱导物时，这些基因不能转录，因为阻遏蛋白以活性状态结合在操纵基因上。当诱导物存在时，阻遏物与之结合，变成失活状态，离开操纵基因，启动子开始转录，起始于 *lacZ* 5′端，终止于 *lacA* 的 3′端。

诱导物如何控制阻遏蛋白的活性呢？阻遏物对于操纵基因有很高的亲和性，在缺乏诱导物时，阻遏物总是结合在操纵基因上，使得邻近的结构基因不能转录。但当诱导物存在时，它和阻遏物结合形成了一个阻遏物复合体，不再和操纵基因结合。

操纵子控制的重要特性是阻遏物的双重性：它既能阻止转录，又能识别小分子诱导物。阻遏物有 2 个结合位点：一个是结合诱导物的，另一个是结合操纵基因的。当诱导物在相应位点结合时，它改变了阻遏蛋白的构象，干扰了另一位点的活性。这种类型的调控叫**变构调控**(allosteric control)。

诱导完成一种**协同调控**(coordinate regulation)：所有的一组基因都一起表达或一起关闭。mRNA 一般总是从 5′开始转录，所以诱导总是导致β－半乳糖苷酶，Lac 透性酶和 Lac 乙酰转移酶按一定序列出现。此多顺反子 mRNA 的共同转录解释了为什么在诱导物的不同条件下，*lacZ*、*Y*、*A* 三个基因的产物总保持同样的当量关系。

诱导触动了"开关"使基因簇表达。诱导物交替变换它们的效应，其他的因子影响了转录和翻译的绝对水平，但三个基因之间的关系事先已被它们的结构所决定。

我们要注意操纵子的潜在特点。*lac* 操纵子含有 *lacZ*，它编码糖代谢所必须的β－半乳糖苷酶；含有的 *lac* 编码透性酶，此酶是负责将底物转达运到细胞中。但操纵子在非诱导状态时，基因尚未表达，也就不存在透性酶。那么诱导物开始怎样进入细胞呢？

其实在细胞中透过酶等总是以最低量存在的，足以供给底物开始进入之需。操纵子有一个**本底水平**(basal level)的表达，即使没有诱导物的存在，它也保持此表达水平

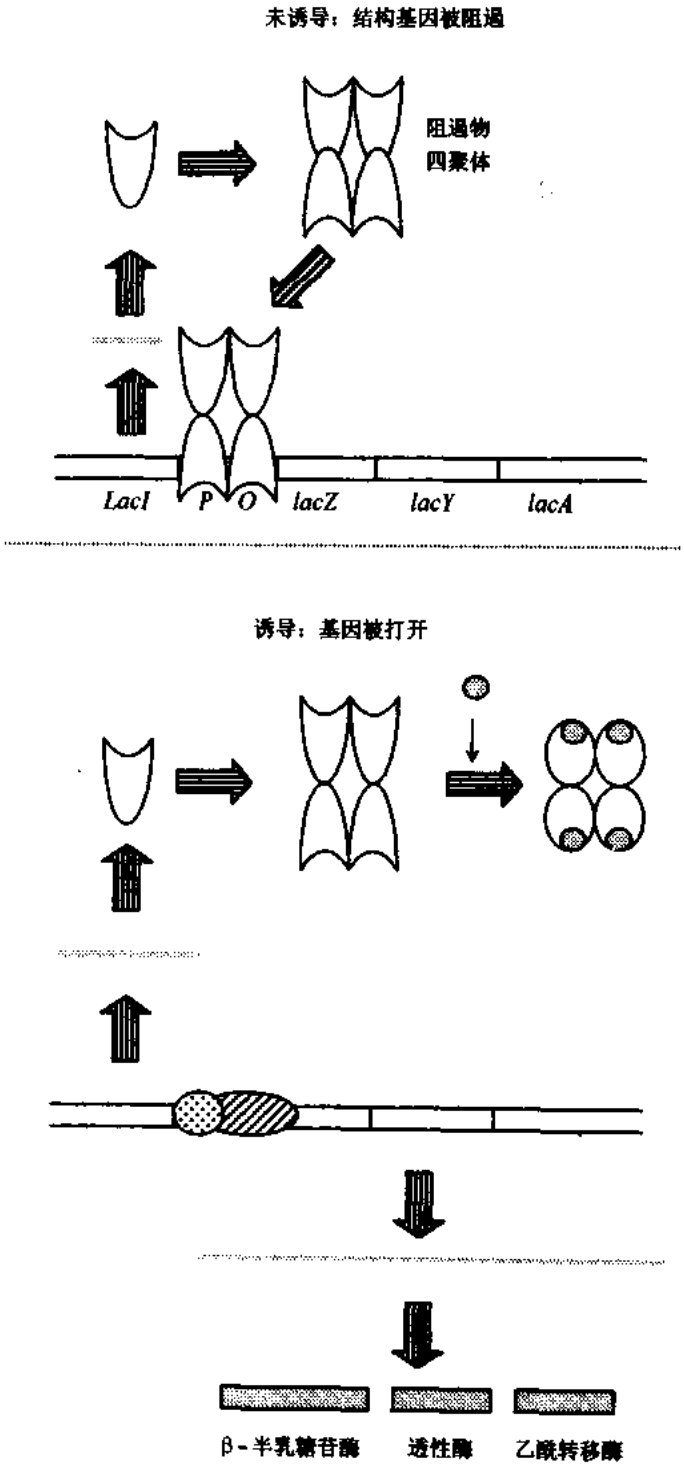


图 16-7　诱导物和阻遏物调节操纵子的开关。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12-6)

(诱导水平的 0.1%),而有的诱导物是通过其他的吸收系统进入细胞的。

3. 操纵基因和调节基因的鉴别

野生型的操纵子以被调节的方式进行表达,调节系统若发生突变可能使表达停止或者在没有诱导物存在时仍然表达。前者称为不**可诱导性**(uninducible)突变;后者对调

节没有反应能力，无论诱导物是否存在都进行表达，故称为**组成型突变**（constitutive mutants）。

lac 操纵子调节系统的成分通过突变可被鉴别出来，它们作用于结构基因的表达。这些成分分为二类：启动子和操纵子作为调节蛋白（RNA 聚酶，阻遏物）的靶序列，通过突变被鉴定为顺式作用调节位点，*lacI* 通过突变被鉴定为具有反式作用的阻遏蛋白基因。

lac 操纵基因是通过组成型突变鉴别出的，称为“O^c”，其分布特点提供了第一个顺式元件的证据，它是有功能的，但本身不编码产物。与 O^c 突变相邻接的结构基因以组成型表达，这是由于突变改变了操纵基因，使阻遏蛋白不能与之结合。这样阻遏蛋白就不能阻止 RNA 聚合酶起始转录。从而使操纵子持续转录。

操纵基因只控制与它相邻接的一些 *lac* 基因。若将第二个 *lac* 操纵子导入细菌的质粒上，它有自己特有的操纵基因。操纵基因互不干扰。因此如果一个操纵子有一个野生型的操纵基因，在通常条件下，它将被阻遏。当第二个操纵子带有 O^c 突变时，它将持续表达。

这些特点表明操纵基因是一个典型的顺式作用位点。操纵基因只控制与其相邻接的基因而不影响存在于细胞中的其他 DNA 上的等位座位。像 O^c 这样的突变称为**顺式显性**（*cis*-dominant）。顺式作用位点中发生突变就不能和相关蛋白相结合，当两个顺式作用位点彼此靠得很近时（如启动子和操纵基因），我们通过互补测验是不能分辨突变发生在一个位点上，只有通过它们对表型的影响来加以区别。顺式显性控制邻接序列的 DNA 位点的特性。如果一个控制位点其功能是作为多顺反子 mRNA 的一部分，它将表现出顺式显性的特点。特别表现在控制位点不能和被它调节的基因相分离。从遗传学的观点来看这些位点和基因是在 DNA 上还是在 RNA 上并不重要。

lacI$^-$ 突变型也可导致持续转录。无论是点突变还是缺失都可产生这样的结果。后者可能是丢失了和 DNA 结合的功能区，因此与诱导物是否存在无关。这种现象是符合负控制系统的。*lac*$^+$ 基因编码一个阻遏蛋白，它可以关闭 *lacZYA* 的转录。阻遏蛋白失去和操纵基因结合能力时，则为组成型突变。转录能在启动子上自由地起始。

当 *lacI*$^-$ 和 *lacI*$^+$ 二者同时存在于同一个细胞时，通过确定二者的关系可以帮助人们得出正确的结论。这只能通过构建**部分二倍体**（partial diploid）来完成的。即一个拷贝的操纵子位于细胞的主染色体上，而另一个放在质粒上，此质粒仅带少量基因，可以独立复制。

在细胞中若既有 *lacI*$^+$ 又有 *lacI*$^-$，则可以正常调节。当除去诱导物时，结构基因又重新被阻遏。这表明 *lacI*$^+$ 可以产生正常的阻遏物，当诱导物不存在时它可以反式阻遏 *lacI ZYA*$^+$ 基因，按遗传学的观点野生型的可诱导性对于组成型突变型是显性的。这是负控制的重要标志。

操纵子非诱导性突变不能得到表达，它们可以分成两种组成型突变：①启动子突变是顺式作用，这种突变阻碍了 RNA 聚合酶与 P_{lac} 的结合，使操纵子不能转录。②*lacI* 突变，阻遏物失去和诱导物结合的能力也会导致和前者相同的现象。这种突变称为 *lacI*^{s}（表 16-1）。

这种反式作用对野生型来说是显性的。阻遏蛋白被保持在对操纵基因的识别和阻碍转录的这种活性状态中，诱导物是否加入对其没有影响。这是由于细胞中突变的阻遏物结合在所有的 *lac* 操纵基因上并阻断转录，同时还不能解离下来，野生型阻遏物的存在对它也毫无影响。

表 16－1　在单倍体和部分二倍体中 β-半乳糖苷酶透过酶的合成

基因型	β-半乳糖苷酶		透过酶	
	诱导	诱导	不诱导	诱导
$I^+Z^+Y^+$	−	+	−	+
$I^-Z^+Y^+$	+	+	+	+
$I^+Z^-Y^+/FI^-Z^+Y^+$	−	+	−	+
$I^-Z^-Y^+/FI^+Z^+Y^+$	−	+	−	+
$I^-Z^-Y^+/FI^-Z^+Y^-$	+	+	+	+
$\nabla(I,Z,Y)/FI^-Z^+Y^+$	+	+	+	+
$O^+Z^+Y^+$	−	+	−	+
$O^+Z^+Y^+/FO^+Z^-Y^+$	−	+	−	+
$O^cZ^+Y^+$	+	+	+	+
$O^+Z^+Y^-/FO^cZ^+Y^+$	+	+	+	+
$O^+Z^+Y^+/FO^cZ^-Y^+$	−	+	+	+
$O^+Z^-Y^+/FO^cZ^+Y^-$	+	+	−	+
$I^{sO}{}^+Z^+Y^+/FO^cZ^+Y^+$	+	+	+	+

∇：表示缺失，“－”表示酶以最高水平存在，“＋”表示酶以最低水平存在

Anthony J. F. Griffiths, et al: An Introduction to Genetic Analysis, 5th ed., 1993, Table 1,2

lacI 突变的特点可以从阻遏蛋白结构得以解释。在阻遏蛋白上具有两种不同类型的结合位点。通过这些结合位点来控制基因的表达以对环境作出反应。DNA 结合位点识别操纵基因，诱导结合位点与小分子诱导物结合。一旦与诱导物作用使其构象发生改变就失去与操纵基因 DNA 结合的能力。通过 *lacI* 突变失去某些活性可以鉴别出阻遏物亚基中的两个结合位点。DNA 结合位点的突变是组成型的（因为阻遏物不能和 DNA 结合来阻断转录）。诱导物结合位点的突变是不可诱导性的（由于诱导物不能改变阻遏物和 DNA 的亲和力）。

阻遏物功能的一个重要的特点是多聚体蛋白。在细胞中阻遏蛋白的亚基随机结合成四聚体。当不同的 *lacI* 等位基因存在时，它们的产物作为亚基结合成异聚四聚体，其特性和同聚四聚体不同。这种亚基之间的作用类型是具有多聚体蛋白的性质，被称为**等位基因间的互补**(interallelic complementation)。

负的互补(negative complementation) 发生在某些阻遏蛋白突变体之间。正如在 $lacI^{-d}$ 与 $lacI^+$ 基因的重组中所见到的一样。此 $lacI^{-d}$ 的突变仅导致阻遏蛋白不能和操纵基因结合。因此它像 $lacI^-$ 等位基因一样，使操纵子呈组成型表达。由于 $lacI^-$ 类型的突变产生的阻遏物没有活性，它相对于野生型基因是隐性的，而“－d”这个符号表示负互补这种突变类型是显性的。这种突变称**反式显性**（*trans*-dominant），也称为**显性失活**(dominant negatives)。

这种显性的原因是由于 $lacI^{-d}$ 等位基因产生一个“坏”的亚基，不仅它本身不能结合操纵基因的 DNA，而且它还通作为四聚体的一部分阻止四聚体中“好”的亚基与 DNA 结合。这就意味着阻遏蛋白四聚体是作为一个总体，而不是单体的简单的集合。这对完成阻遏来说是很必要的。在体外将“好”的亚基和“坏”的亚基混合起来也不能产生正常的功能。

$lacI^{-d}$ 的突变发生在阻遏蛋白的 DNA 结合位点（图 16－8），这就可以解释混合的四聚体可以阻止与操纵基因的结合。结合位点数目的减少使四聚体和操纵基因的亲

和力减少。*lacI* 基因作图表明(图 16－8) *lacI*$^{-d}$ 突变位点是成束排列在基因的左末端对应于蛋白产物 N 末端 DNA 结合位点。*lacI*$^{-}$隐性突变发生在此位点以外的任何区域。但可间接地影响到阻遏物和 DNA 的结合。

lacI^{s} 是不可诱导性突变,它不能对诱导物作出反应。这可能由于阻遏蛋白失去了诱导物结合位点,或者不能将它们的作用传递到 DNA－结合位点。在图 16－8 中可以看出 *lacI*^{s} 突变位点很有规律的沿着基因成束间隔排列。这些间隔可能存在着肽链的改变。

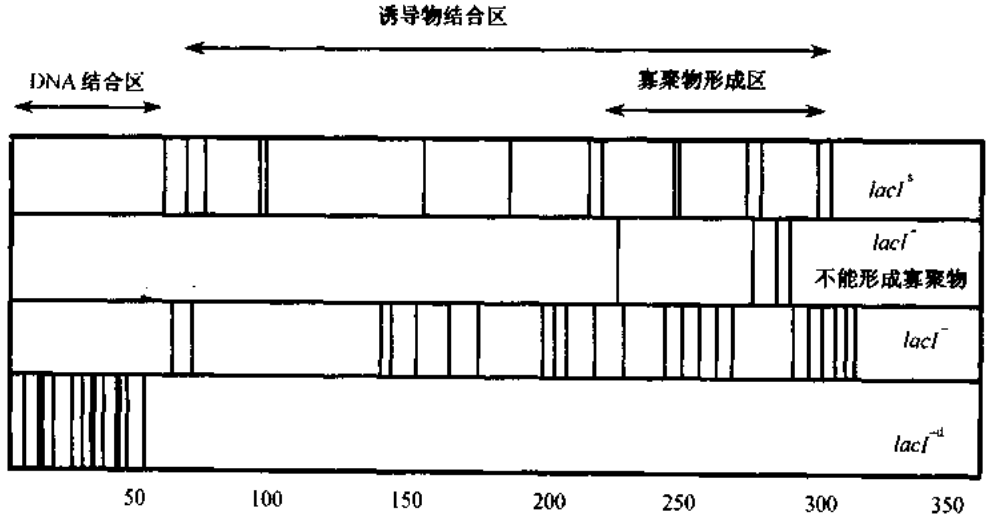


图 16－8 *lacI* 基因不同功能区的突变。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－11)

4. 阻遏蛋白和操纵基因的结合与解离

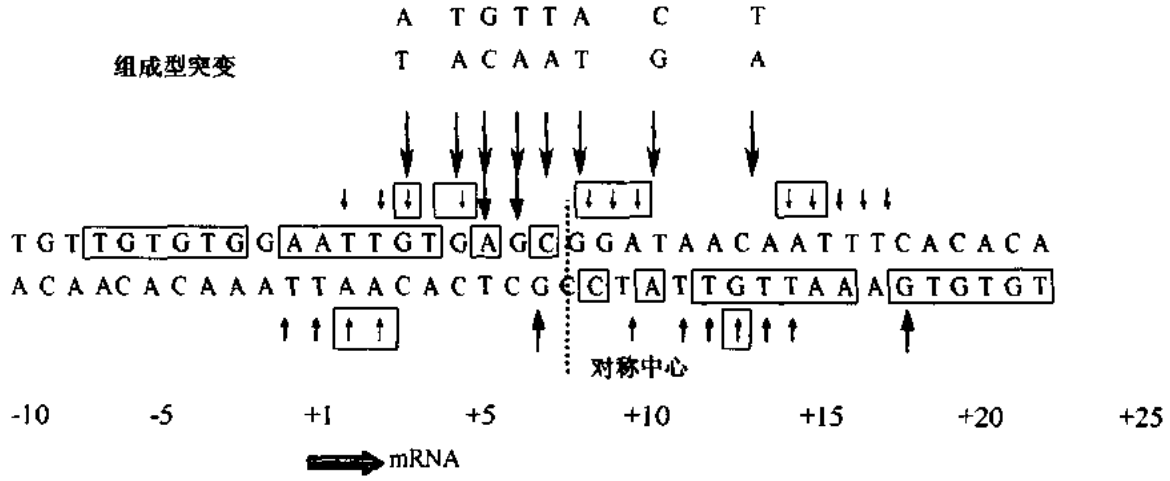
经分离纯化的阻遏蛋白的成分可以和安慰诱导物 IPTG 结合的。(由于在细胞中阻遏物的总量很少，为了获得足够的阻遏物，必须要使用上升突变的启动子来增加 *lacI* 转录。然后将它们构建在多拷贝的质粒中,这样可以使产物增加 100～1000 倍)。

阻遏蛋白结合的双链 DNA 区含有野生型的 *lac* 操纵基因的序列。但阻遏物不能结合发生 *O*^{c} 突变的 DNA 序列。在体外实验中,加入 IPTG 时阻遏蛋白就可以从 *Olac* DNA 上解离下来。根据在体外阻遏蛋白与 O_{lac} DNA 之间的相互作用所显示的特点可以推测体内发生的情况。

阻遏物怎样识别操纵基因特异的 DNA 序列的呢? 阻遏物和操纵基因之间的相互作用常作为 DNA 特异序列综合反应的范例。操纵基因的共同的特点就是有很多细菌调节蛋白的识别位点,即回文序列(图 16－9)。

人们采用分析 RNA 聚合酶与启动子相互作用的方法来研究阻遏物和操纵基因的连接点。两侧中任何一侧发生缺失都可以用来分析此区域的末端位点。组成型点突变所鉴定出的单个碱基对是很重要的位点。在 DNA 和阻遏物的结合实验中,将与阻遏物相结合的 DNA 和未结合的 DNA 分别进行甲基化，比较它们对甲基化的敏感性，或者分别用 UV 照射,比较产生变异碱基的情况来表明 DNA 是否受到蛋白的保护。

阻遏物所结合的 DNA 对称区域因受到保护而不被核酸酶所降解。此区域从－5 到＋21 共 26bp,通过组成型突变已将此区域鉴别出来,它是很小的一段 DNA,从＋5 到＋17 这 13bp 是中心区,在中心区有 8 个碱基对的替代可导致组成型的突变。前面已介



图例:受到阻遏物保护的 Pu 抗甲基化;↓ 此处的 P 通过阻遏物增强了甲基化;↓ 可与阻遏蛋白连接的 T

图 16－9　lac 操纵位点的对称顺序。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－12)

绍启动子的突变也在这几个位点上。在一个大的区域含有的少量特异的碱基对常常负责和蛋白的特异性结合。碱基的增强和保护模式表明紧密相连的 DNA 区域中对称排列特点。图 16－9 表明从 +1 到 +22 的区域中所有的 T 除 2 个以外都可以与阻遏物交联。甲基化实验表明在 +3 与 +9 之间 Pu(A 和 G)占的比例很大,它们受到阻遏结合的强有力保护。只有少数碱基对加强了对甲基化的敏感性,可能由于在阻遏物和操纵基因复合物中产生了"疏水囊"。

+1 ~ +6 这段序列和 +21 ~ +16 的序列存在着潜在的对称性。和对称轴紧密相连的区域并不是对称的,组成型突变的模式表明了这一点。+5/+7 及 +8/+14 是对称的,但紧靠着对称轴的位点不是对称的。当阻遏物位于对称物附近时,它是以边远位点交连方式骑在对称序列上。

lac 操纵基因的反向重复序列的两侧功能并不相同。突变位点的分布表明操纵基因的左侧对于损伤是较为敏感的。它含有 6 个点突变,而右侧只有 2 个点突变位点。左右突变位点的发生也不是对称的,左边的突变作用较大。对称与阻遏物——操纵基因的连接紧密相关,看来阻碍物与左侧结合更为紧密。通过突变来增加对称性,即让左侧的与右侧的反向重复更为完全,可以使操纵基因和阻遏物的亲和性增加 10 倍左右。典型的操纵基因是含对称重复序列的右半个。

DNA 序列的对称性是否反映了蛋白质的对称性呢? 由于阻遏蛋白是由相同亚基组成的四聚体,每个亚基必须都具有相同的 DNA－结合位点。操纵基因的每个反向重复和阻遏物的二聚体相结合。就像很多其他的细菌调节物一样,*lac* 阻遏物以一个短的 α 螺旋区与 DNA 相结合。此 α－螺旋的序列决定了与 DNA 的特异结合。

据推测,阻遏物四聚体与操纵基因紧密结合。诱导物进入细胞,减少了阻遏物和操纵基因之间的亲和性,但诱导物怎样使阻遏物从操纵基因上释放出来呢?

各种诱导物在体外可导致阻遏物和操纵基因亲和性降低。这表明诱导是由于操纵基因和阻遏物之间的亲合力降低。阻遏物从操纵基因上分离下来的速率是较低的,在体外左半个操纵基因要 10～20 分钟,但当加入 IPTG 时,操纵基因和阻遏物多聚体的稳定性立即降低,显然其半衰期明显减少。

阻遏物作用有两种模型：

(1)平衡模型(间接作用)：结合在 DNA 上的阻遏蛋白和游离的阻遏蛋白迅速达到平衡；诱导物和游离的阻遏蛋白相结合。这样阻止其与 DNA 的重新结合，部分结合在 DNA 上的阻遏蛋白解离下来，以达到新的平衡。

(2)解离模型(直接作用)：阻遏蛋白从操纵基因上解离下来的速率非常低(在缺乏诱导物时检测的结果)以至不能和平衡模型相吻合。这就意味着诱导物必须直接和结合在操纵基因上的阻遏蛋白相互作用。可能通过改变其构象使其离开操纵基因(图 16－10)。

用高浓度蛋白质进行甲基化保护或刺激作用可用来研究阻遏物－IPTG 复合物与操纵基因的结合。阻遏物－IPTG 复合物对于操纵基因的亲和性很低，对这种亲和性进行大剂量补偿。这种复合物与 DNA 精确接触的方式与游离的阻遏物相同。有些突变的

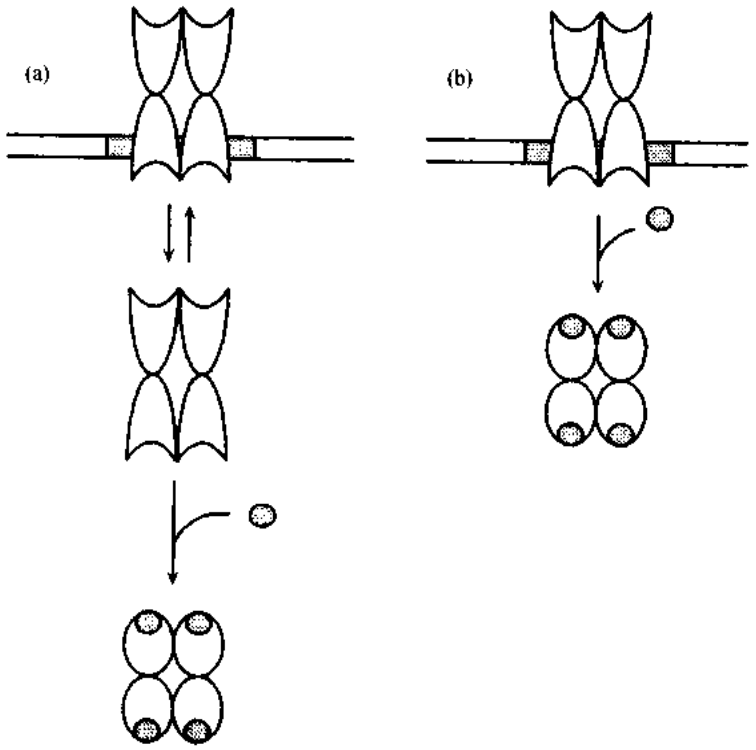


图 16－10　诱导物作用模型。(a)平衡模型，诱导物与游离的阻遏物结合，打破了其游离与结合形式之间的平衡；(b)解离模型，诱导物直接与结合在操纵位点上的阻遏物结合，使其释放。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 12－13)

阻遏蛋白和 DNA 操纵基因的亲和能力是增加的，但它们的接触方式也相同。各种变异的阻遏蛋白，尽管它们的亲和常数相差 7 个数量级，但都以相同的方式与 DNA 相接触。阻遏蛋白与 DNA 亲和能力的改变一定是受到了蛋白构象的影响。而不可能是一、两个键的断裂所致。

对在体外分离的阻遏蛋白的研究表明了它具有 DNA 结合位点，此位点与 DNA 的结合能力受到分子结构等因素的影响。当用胰蛋白酶来处理阻遏蛋白时，可在第 59 个氨基酸处切开阻遏蛋白，此蛋白的 C－端的片段含 60～360 氨基酸，称为**抗胰蛋白酶核心**(trypsin－resistant core)。它具有聚集成四聚体以及和诱导物结合的能力，但不能

与操纵基因相结合。N－端 1～59 氨基酸的片段称为**短的头部片段**(short headpiece)。短的和长的头部片段保留与 DNA 的结合能力，当操纵基因存在时头部片段和完整的阻遏物一样能与 DNA 结合(但要比完整的阻遏蛋白结合得慢)。头部片段与操纵基因结合能力表明其结构相对独立于蛋白的其他部分。

这与下面的结构模型相吻合。蛋白体 N 端伸出的 50 个氨基酸的**臂或突起**(arm or protrusiton)作为 DNA 的结合位点。**臂由交连区**(长 50～80 氨基酸)连接；其余的 81～360 氨基酸的 C－端是负责四聚体的缔合以及和诱导物相结合(图 16－11)。阻遏物单体的 N 端由 2 个 α 螺旋构成，它们是 DNA 结合基序，称为**螺旋转角螺旋**(helix－turn－helix，HTH)，它在 DNA 的大沟与特定的碱基结合。这一区域是由一个 α 螺旋构成的铰链区与 C－端的主体蛋白相连，主体蛋白称为**核心区**(core)。核心区由两个相似的结构区组成，分别称为**核心区域 1 和 2**。每一个核心区都有 6 个平行的 β 折叠，其两侧是 α 螺旋。诱导物就结合在这两个区域之间的裂缝中。在 C－末端有一个 α 螺旋，在 α 螺旋上有 2 组 Leu 的重复，每组 7 个 Leu，由此区形成阻遏物四聚体。

通过各种实验我们知道了诱导物的结合可导致阻遏蛋白产生重要的构象改变，两个分子的诱导物和四聚体阻遏蛋白相结合就足以使诱导物释放。单独的头部片段的构象改变表明构象变化的模型。当诱导物结合时从 DNA 上释放出阻遏物(图 16－12)，

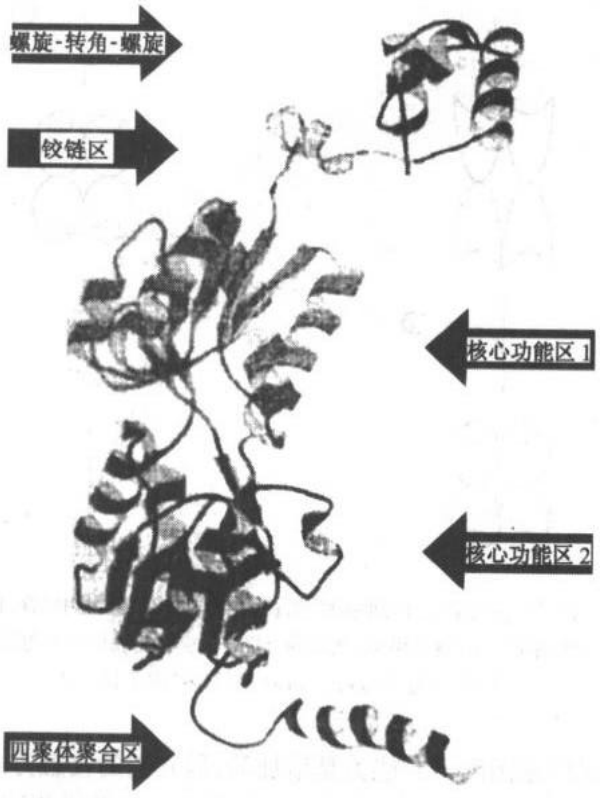


图 16－11　Lac 阻遏蛋白单体的结构。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig. 12－14)

DNA 结合位点构象从完全适合于 DNA 序列的状态变成为另一种状态，交连的角度改变，从而不能和 DNA 紧密结合。

阻遏蛋白结合在操纵子上对 DNA 聚合酶结合启动子有什么影响呢？这两种蛋白可以同时结合在 DNA 上，而阻遏蛋白的结合实际上增强了 RNA 聚合酶的结合。但酶

图 16-12 诱导物改变了阻遏蛋白二聚体头部片断核心功能区的构象，使其不再和操纵基因的 DNA 结合。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig. 12-16)

的结合阻止了转录的起始。

RNA 聚合酶与 *lac* 启动子的结合是保持一定的量：$1.9\times10^{7}M^{-1}$，而阻遏物的存在增加了此结合的量，使其高出 2 个数量级，为 $2.5\times10^{9}M^{-1}$，打破了 RNA 聚合酶和 DNA 结合的平衡常数 K_B，阻遏蛋白使 RNA 聚合酶与 *lac* 启动子的相互作用由弱到强，从而有效地改变闭合二元复合体的形成。

较高的 K_B 值表示当操纵基因被阻遏蛋白占据时，启动子与 RNA pol 结合能力约提高 100 多倍，(如果 RNA pol 的浓度是一定的话)，并允许 RNA pol 在阻遏物与 DNA 结合的同时与启动子结合。有可能一经诱导，结合在启动子上的 RNA pol，便立即转录。

实际上阻遏物导致 RNA 聚合酶在启动子上的贮存，RNA 聚合酶－阻遏蛋白 DNA 复合体堵塞在闭合复合体阶段，当诱导物加入时阻遏蛋白被释放，闭合复合体变成为开放复合体，立即起始转录。阻遏物加速了诱导的过程。

这个模型是否可以应用于别的系统呢？RNA pol 阻遏物和启动子操纵基因的相互作用在各系统中都是明显的，由操纵基因并不都是和启动子重叠(图 16-13)，如 λ 噬菌体操纵基因位于启动子的上游区(见本章)，因而阻遏物在别的系统中可能不是以这样的方法和 RNA pol 起作用。

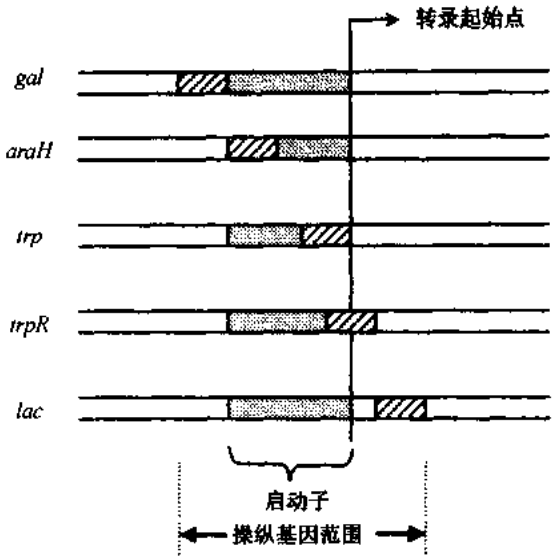


图 16 - 13　操纵基因可位于启动子不同的相对位点。(仿 B.Lewin:Genes Ⅵ,1997, Fig.12 - 20)

5. 蛋白和 DNA 相互作用的特异性

调节蛋白对于特异位点具有较高的亲和性，而对于其他随机位点具有较低的亲和性。阻遏蛋白四聚体正是如此，有大量的**低亲和性位点**和少数的**高亲和性位点**。在 *E.coli* 基因组中阻遏蛋白仅有一个高亲和性位点，那就是操纵基因。而其他的地方都是低亲和性位点。在基因组中每一个碱基对都可以作为一个新的低亲和性位点(延着基因组移动一个碱基，而与操纵基因的相位不同，就可以产生一个低亲和性位点。因此 *E.coli* 具有 4.2×10^6 低亲和性位点。大量的低亲和位点意味着在缺乏特异结合位点时几乎所有的阻遏蛋白都结合到 DNA 上，而无游离阻遏蛋白的存在。

我们可以通过下列的平衡来描叙与阻遏蛋白与 DNA 的结合。

$$K_A = \frac{[\text{阻遏蛋白} - \text{DNA}]}{[\text{游离阻遏蛋白}][\text{DNA}]}$$

[阻遏蛋白 - DNA] 表示结合于 DNA 的阻遏蛋白的浓度；[游离阻遏蛋白] 表示游离阻遏蛋白的浓度。而[DNA]表示非特异性结合位点的含量。

总的阻遏蛋白中有多少是游离的呢？我们通过下列公式来看看它们的分布。

$$\frac{[\text{阻遏蛋白}]}{[\text{游离阻遏蛋白} - \text{DNA}]} = \frac{1}{K_A \cdot [\text{KDA}]}$$

(1)非特异的平衡结合常数是：　$K_A = 2\times10^6 M^{-1}$

(2)非特异结合位点的含量是每 10^{-15} 升细菌中有 4×10^6 个位点。相当于[DNA] $= 7\times10^{-3}$mol/L (这是很高的含量)

游离阻遏蛋白/结合的阻遏蛋白 $= 1/(2\times10^6\times7\times10^{-3}) = 10^{-4}$

因此几乎所有的阻遏物都结合在 DNA 上，仅有 0.01%的阻遏蛋白是游离的。由于在每个细胞中约有 10 个分子的阻遏蛋白，也就是说没有游离的阻遏蛋白，这就意味着阻遏蛋白和操纵子的相互作用，涉及到阻遏蛋白以不同的亲和力结合在 DNA 上。与大

量低亲和位点相比，只有单个高亲和位点。

表 16－2 比较 *lac* 阻遏物和操纵基因结合以及阻遏物和一般 DNA 结合二者的平衡常数。从这些常数中我们可以推理如何区分结合操纵基因的和结合其他 DNA 的阻遏蛋白，而且可以从中了解到当诱导物导致阻遏物以操纵基因上解离下来时，究竟发生了什么事。

阻遏物与操纵基因的结合要比与相同长度的随机 DNA 顺序的结合亲合力约大 10^7 左右。这样操纵基因由单个高亲和性位点构成，而它的竞争性要比任何低亲和位点

表 16－2 Lac 阻遏物结合操纵基因的强度和特异性(所有的平衡常数是 M^{-1})

DNA	阻遏物	阻遏物＋诱导物
操纵基因	2×10^{13}	2×10^{10}
其他 DNA	2×10^{6}	2×10^{6}
特异性	10^7	10^4

B. Lewin: Genes Ⅵ. 1977, Table 12.1

强 10^7 倍。

利用特异性我们可计算出阻遏物在操纵基因以及其他随机位点之间的分布情况，并能以操纵基因的“占据”这一术语来表达。如果每个细胞中有 10 个 *lac* 阻遏蛋白分子对于操纵基因的特异性为 10^7，即操纵基因大部分时间（96%）都被阻遏蛋白所结合。*lac* 阻遏物－操纵子相互作用的特异性，有 2 个特点：

(1) 当诱导物结合在阻遏物上，阻遏物对操纵基因的亲和能力降低 10^{-3} 倍，而与普通的 DNA 顺序的亲和能力将保持不变。因此和诱导物结合后操纵基因的特异性仅为 10^4。此不足以和过量的低亲和力(4.2×10^6)位点来竞争阻遏物。因此在这种情况下仅有 30%的操纵基因和阻遏物结合。

(2) 当操纵基因发生突变，使其和阻遏物的亲和性下降约 20～30 倍，这种突变必然影响到结构。在基因组中，操纵基因的突变体可能竞争不过占优势的随机位点。如果阻遏物的特异性减少 10 倍，那么操纵基因的占据将要下降到 50%左右。

在不被诱导的细胞中，阻遏物四聚体通常结合在操纵基因上。图 16－14 表明所有的或几乎所有余下的四聚体都随机地结合在 DNA 其他的区域。在细胞中很少或没有四聚体游离于细胞中。加入诱导物就解除阻遏物特异地结合在操纵基因上的能力，使它们从操纵基因释放出来，而移到低亲和力的随机位点上。这样，在一个被诱导的细胞中，那些阻遏物四聚体被储存在随机 DNA 位点上。在未诱导的细胞中阻遏物结合在操纵基因上，余下的四聚体分子被结合在非特异位点上。因此诱导的作用是改变四聚体的分布，而不是产生游离的阻遏物。

当诱导物消失后，阻遏蛋白又很快地恢复了它特异结合操纵基因的能力。这必然涉及到它从 DNA 非特异的“贮存”位点上的移动。那么迅速移动的机制是什么呢？这种迅速结合操纵基因的能力并不和解离与重新结合多重循环的时间相一致，表明可以排除发现操纵基因的随机碰撞机制，而可能沿着 DNA 从随机位点直接移向操纵基因。这种方式和前面所涉及的 RNA Pol 寻找其启动子的方式相同(图 13－14)，也可能通过成环后直接取代，四聚体以随机位点直接移到操纵基因。每个四聚体存在的结合位点是 4 个，而每次结合 DNA 时只使用了其中的 2 个，因此余下还有 2 个结合位点，这对取代反应是有利的。高亲和力操纵基因与很多低亲和力位点的竞争，使阻遏物处于“左右为难”的境地。在阻遏的条件下，操纵基因必须有高度的特异性，即高亲和性；但在诱导

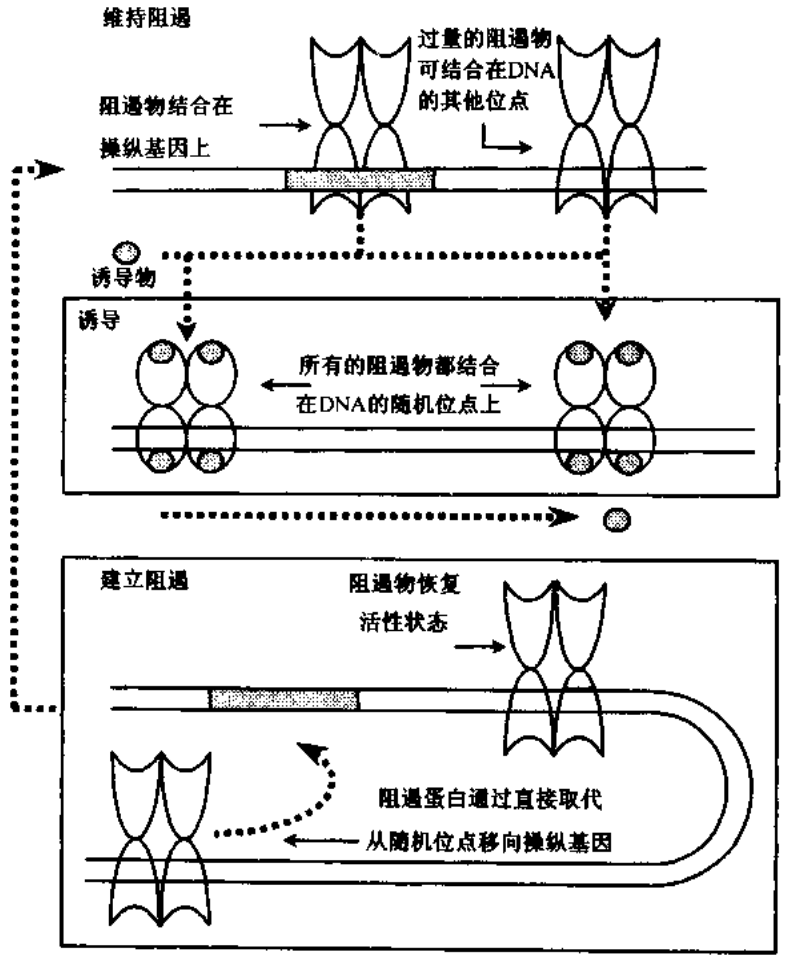


图 16－14　在细胞中所有的阻遏蛋白都结合在 DNA 上。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－18）

时对这种特异性必须减低，即亲和力减低。例如假设每个细胞中有 1000 个分子的阻遏蛋白，在阻遏条件下仅有 0.04% 的操纵基因是游离的。因此要将完全阻遏的状况变成为部分阻遏，细胞内阻遏蛋白的合成数目要服从于供需平衡。

lac 操纵子的表达水平在诱导和阻遏之间的差别实际上是 10^3 倍。从这一角度来说，即使是缺乏诱导物，也存在本底水平的表达，约为诱导时表达水平的 0.1% 左右。若阻遏蛋白存在较多，那么表达水平就下降，若其存在较少，表达水平就上升。若每个细胞中阻遏蛋白要少于 10 个就不可能建立起严密的阻遏。若阻遏蛋白太多，诱导操纵子变得较困难。

三、阻遏能发生在多位点上

lac 阻遏蛋白只作用在 *lacZ*、*Y*、*A* 基因簇的操纵基因上，然而其他的阻遏物通过结合一个以上的操纵基因来控制分散的结构基因。例如 *trp* 阻遏物控制三个不连锁的基因。

在基因簇 *trpEDBCA* 中的操纵基因控制催化分支酸合成色氨酸所需的一系列酶的合成（图 16－15）。

aroH 基因是芳香氨基酸生物合成途径中编码 3 个酶中的一种酶的基因，它催化合成反应的起始。在另一位点上的操纵基因可以控制这个基因。

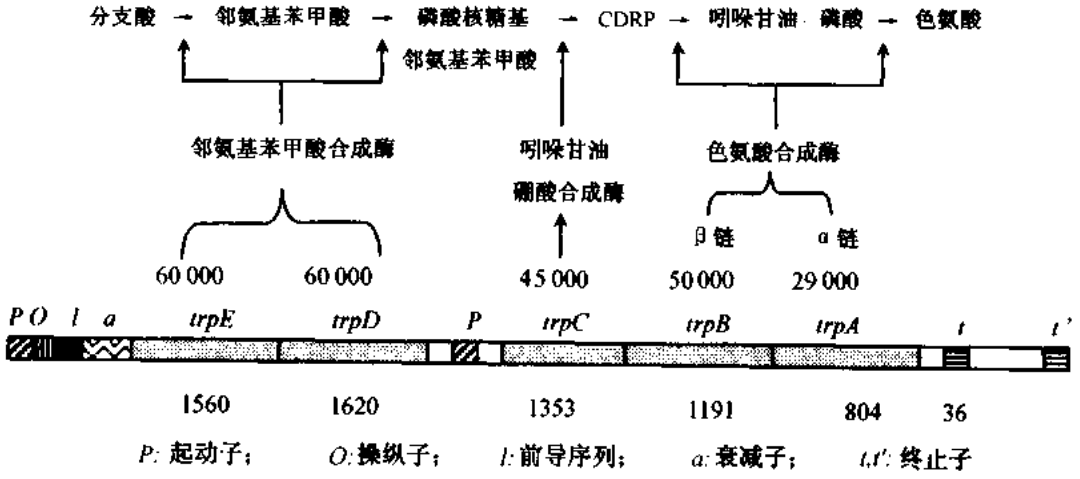


图 16－15　*E.coli trpO* 的结构及其产物所催化的色氨酸合成反应。(仿 B.Lewin: Genes Ⅵ, 1990, Fig. 15－2)

trpR 调节子基因可被它自己的产物 *trp* 阻遏物所阻遏，这样阻遏蛋白的作用可以减少它本身的合成，这是自我控制的一个例子。这种情况在调节基因中是十分普遍的，它可能是正调控，也可能是负调控。

trp 阻遏物作用的 3 个位点的每一个都存在 21bp 长的操纵基因序列。这些序列的保守性如图 16－16 所示。每个操纵基因的结构可以估计是二侧对称的。三个操纵基因的保守序列含有 *trp* 阻遏物的接触位点。每个位点都具有一个拷贝的特异序列结合，它可被阻遏物所识别，正如每个启动子具有与其他启动子一致的序列一样。这可以解释为什么一种阻遏物可以作用几个位点。

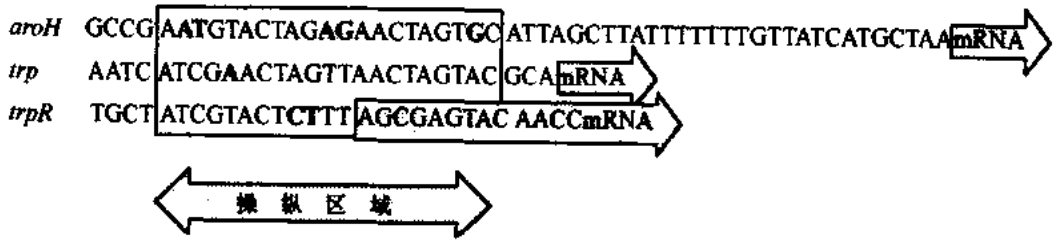


图 16－16　*trp* 阻遏物识别三个位点上的操纵区。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 12－19)

图 16－13 表明操纵基因和启动子的相对关系。由 TrpR 识别的操纵基因的分布重要特点是它们位于启动子不同的相关位点上。在 *trpR* 中，操纵基因位于 －12 和 +9 之间；在 *trp* 操纵子中是位于 －23～－3；而在 *aroH* 中是位于更上游处，在 －49～－29 之间；但在 *lac* 操纵子中，是位于启动子的下游。与此相反在 *gal* 操纵子中，操纵基因是位于启动子的上游。阻遏物作用在操纵基因上，它们在不同的靶启动子中位置不同，表明它们具有不同的阻遏方式，但共同的特点是 RNA 聚合酶在启动子上受阻而不能起始转录。

四、分解代谢的阻遏涉及到启动子的正调控

只要我们一提到启动子，人们立即会想到那是一段能与 RNA Pol 结合的 DNA 序列，结合后就可以转录。其实 RNA Pol 在有的启动子上结合并不能起始转录。它还需要一种附属蛋白的帮助。这种蛋白属于正调控因子。因为它们的存在是启动转录单位所必须的。已知各种正调控蛋白有的是由噬菌体编码的，有的存在于宿主的细胞中。它们的激活方式之一是激活物填补启动子中的空缺处。例如 – 35 或 – 10 处的保守序列。

大部分激活因子都是蛋白质，它们控制 *E. coli* 操纵子靶位点的活性，从而对碳源的状态作出反应。当葡萄糖作为一种有效的碳源时，它优先被利用。这样当 *E. coli* 的培养基中同时存在葡萄糖和乳糖时，细菌代谢葡萄糖而不使用乳糖。

这种选择是通过阻止各种操纵子的表达来完成的，包括乳糖，半乳糖和阿拉伯糖操纵子。这种作用称为**分解代谢的阻遏**(catabolite repression)。它通过抑制有关酶的操纵子表达而优先利用葡萄糖。

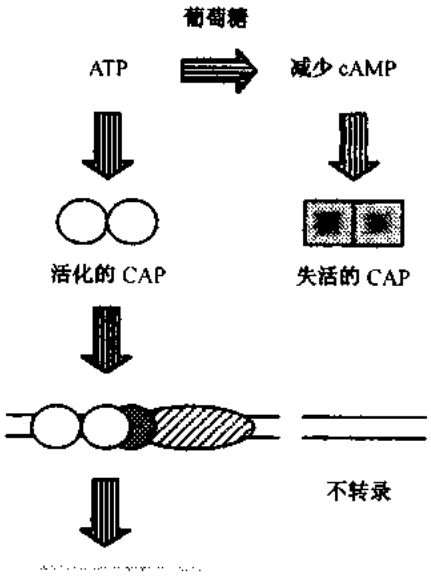


图 16 – 17　通过减少 cAMP 的水平葡萄糖导致了分解代谢的阻遏。(仿 B. Lewin: Genes Ⅵ, 1997, Fig12 – 23)

分解代谢的阻遏是通过葡萄糖的存在可以减少细胞中 cAMP 的水平。分解代谢 – 调节操纵子的表达要涉及到 cAMP 的浓度，调节作用如图 16 – 17 所示。

cAMP 由**腺苷环化酶**(adenylate cyclase)催化合成，此反应是以 ATP 为底物，以磷酸二酯键将 3′ – 5′相连接，产生了环状 AMP 结构(图 16 – 18)。若编码腺苷环化酶的基因发生突变(cya^-)就不出现分解代谢的阻遏。

在 cap^- 中鉴别出 **CAP**(catabolite gene activator protein)，这个调节蛋白是直接作用在靶操纵子上的。它是一个正调控因子，启动子的转录起始需要它的存在。这种蛋白只有在 cAMP 存在时才有活性。cAMP 作为一种小分子的诱导物(图 16 – 18, 19)，若 cAMP 的水平减少就会使这种蛋白不能结合在控制区，从而阻止 RNA pol 起始转录。因此葡萄糖的作用是减少 cAMP 水平，剥夺了操纵子表达时所需要的控制因子。

CAP 因子结合于 DNA 上，且 cAMP – CAP – DNA 的复合物可以在多个启动子上分离得到。此因子由两个相同的亚基 (22 500kDa) 组成为二聚体。它可被单个分子的 cAMP 所激活。CAP 的单体保持在 DNA 结合区和转录激活区。

CAP 二聚体结合在启动子上约 22bp 的位点。此位点由一保守的回文序列组成(图 16 – 19)。若发生突变将会阻止 CAP 的作用，通常 $\frac{TGTGA}{ACACT}$ 是提供识别的一种潜在的元件。CAP 能紧密地结合在这两个反转重复的五聚体上，每个亚基结合一个五聚体 DNA。很多的结合位点缺少了第二个五聚体，那么第二个亚基就必须结合在一个不同

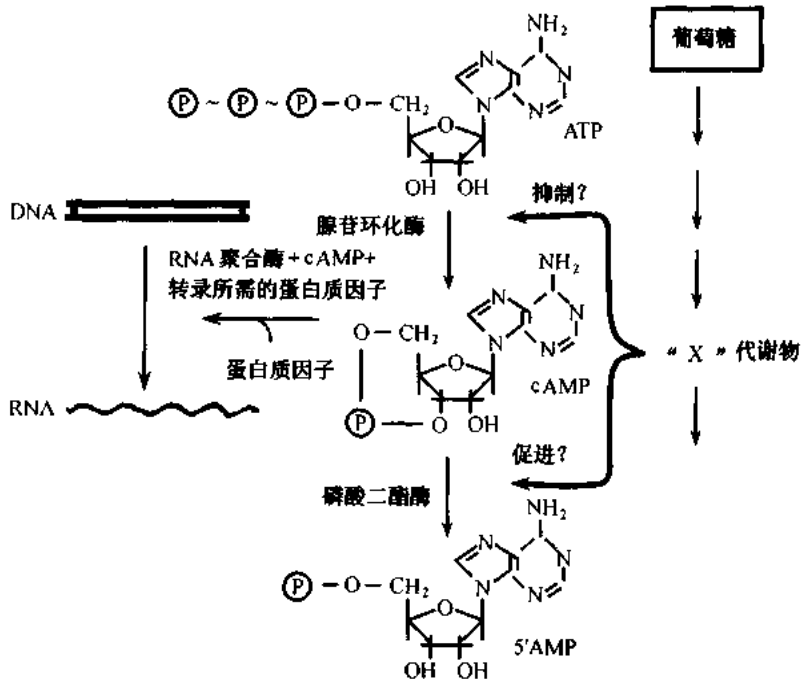


图 16－18　通过 cAMP 控制代谢敏感性转录。(仿 J.D.Watson *et al*.: Molecular Biology of the Gene, 4th ed., 1987, Fig. 16－14)

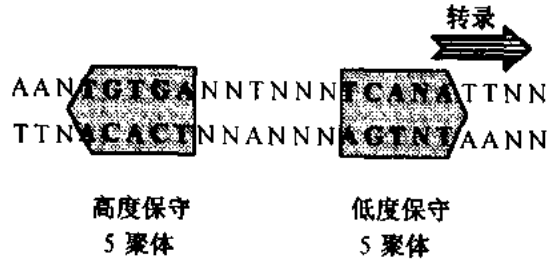


图 16－19　CAP 结合位点的保守序列。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 12－2)

的序列上，CAP 结合效率的等级可以有助于解释为什么不同的基因在体内被不同水平的 cAMP 所激活。

CAP 的作用的一个特点，是在不同的操纵子中其结合位点位于相对于起始点的不同位置。即 TGTGA 五聚体可以位于不同的位置(图 16－20)。

(1) CAP－结合位点位于启动子中，如 *gal* 操纵子的 CAP 结合位点中心在－41，从－50 延伸到－23，看来只有单个的 CAP 二聚体结合。由于 CAP 结合位点将延伸到 RNA 聚合酶所保护的区域，可能十分紧密地和 RNA 聚合酶接触。

(2) CAP－结合位点与启动子邻接，如 *lac* 操纵子，其 CAP 结合区的中心在－61，从－72 延伸到－52，它可能和 2 个 CAP 二聚体结合。这种结合方式与 CAP 大部分存在于 DNA 的一个面上的情况相一致。RNA 聚合酶也在 DNA 的同一个面上结合。这两种蛋白在结合位点上可能会互相接触。

(3) 在另一种操纵子中，CAP 结合位点位于启动子的上游，如 *ara* 操纵子。单个 CAP 的结合位点离转录起始点更远，在－107～－78，这样 CAP 不能和 RNA pol 接触，因另一个调节蛋白要结合在 CAP 结合位点与 RNA pol 位点之间的区域(图 16－20)。

启动子内在的效率是依赖于 CAP，不依赖于启动子的－35 序列，而也有的缺少了

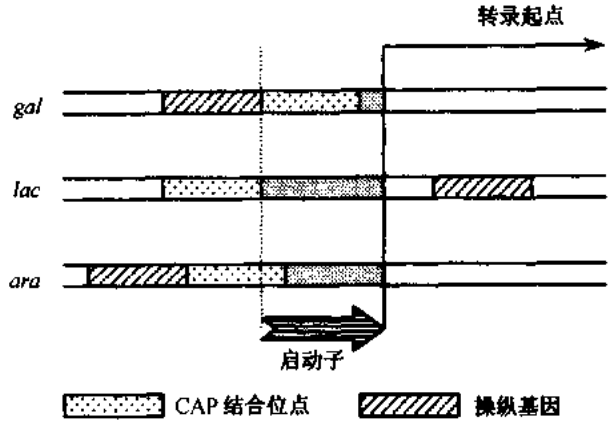


图 16 - 20　CAP 蛋白可结合于相对于启动子的不同位点。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12 - 25)

- 10 顺序。另一种观点认为,如果启动子具有 - 35 和 - 10 区域,它可以独立地与 RNA pol 相互作用,CAP 很难控制其效率。

原则上 CAP 有两种方法来激活转录:(1)它可能直接和 RNA pol 相互作用;(2)或者它作用于 DNA,以某种方式改变其结构,从而帮助 RNA pol 结合。其实 CAP 具有作用 RNA pol 和 DNA 两方面的功能。

RNA pol 能由它们的亚基重新聚集而成,而当 α 亚基 C 末端有缺失时转录仍然正常,但失去了被 CAP 激活的能力。这表明 CAP 直接作用在 α - 亚基,刺激了 RNA pol。用 CAP 二聚体的一个亚基来进行实验,这个亚基具有功能正常的转录激活区,当此 CAP 结合到 *lac* 启动子上时,只有转录起始点附近的亚基的激活区是必要的。这是由于它与 RNA pol 接触。这就提供了一种解释:结合位点并不依赖一定的方向:CAP 的二聚体结构保证了有一个亚基可以为 RNA pol 提供激活区。

作为一般原理,当两个或更多的蛋白结合在 DNA 上时,转录复合体的形成要借助于蛋白与蛋白之间的相互作用。两个蛋白之间适度的相互作用就足以增加 RNA pol 启动子的结合能力。

与结合的 RNA pol 的作用依赖于两种蛋白的相对位置。当 CAP 结合在启动子中时(如 *gal* 操纵子),它增加了由闭合复合体转化为开放复合体的速率。当 CAP 结合在启动子相邻位置(如 *lac* 操纵子)它的优势作用增加了起始结合的速率,形成了一个闭合复合体。这表明 CAP 和 RNA pol 之间的相互作用的影响,取决于蛋白质在单个启动子上的几何学(*ara* 启动子的相互作用可能涉及到更多蛋白的相互作用)。

在大部分启动子上的 CAP 结合位点,都和 *gal* 或 *lac* 操纵子相似。当距离改变,CAP 激活转录的能力降低,这反映了 CAP 和 RNA 必须结合在 DNA 的同一个面上。

CAP - DNA 复合体的结构是很有趣的:DNA 有一个弯曲。当结合时 CAP 可能使 DNA 双螺旋变得弯曲,各种调节蛋白都可能导致轴变弯。

图 16 - 21 表明能用来量度 DNA 弯曲的程度及位置的一种技术。先制备一个二聚体靶顺序(有两个蛋白结合位点),然后用不同的限制酶切,产生一套只含一个结合位点且位置不同的 DNA 片段。由于弯曲的位置不同使得各片段在电泳胶上的迁移率也不同(若不弯曲的话各片段的泳动率是相同的)。当结合位点位于片段的中央时,运动所受的阻力最大,泳动率最低;当结合位点位于片段的末端时,运动所受的阻力最小,泳动率最高。不同的酶切片段,通过泳动率的结果可进行分析。最快的泳动率在曲线的

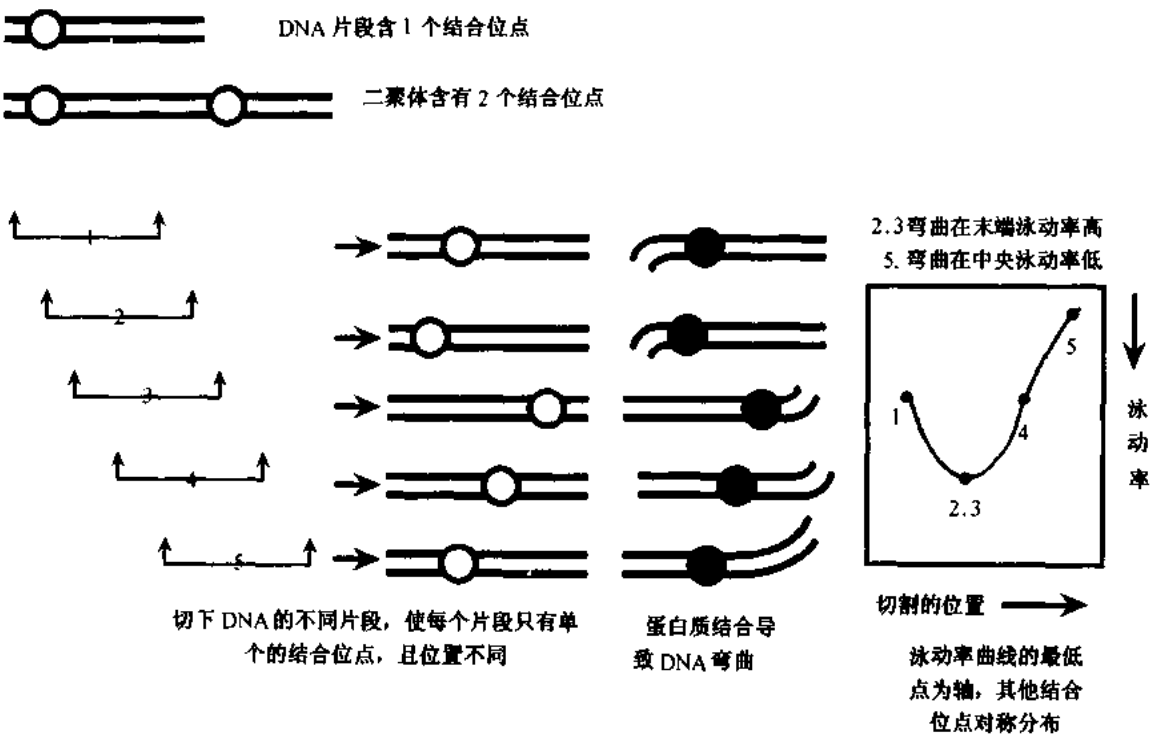


图 16－21 用电泳来分析结合位点。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－26)

最低点，这表明酶切位点就在结合位点的附近。

CAP 和 *lac* 操纵子的相互作用位置在对称轴的中心。DNA 弯曲得很厉害（> 90°）（如图 16－22 所示）。当 CAP 与 DNA 结合时，使双螺旋的结构发生了戏剧性的改变。弯曲对转录可能有某些直接的作用，但也可能是使 CAP 在启动子上可以接触到 RNA pol。

无论 CAP 在各种启动子上激活转录机制是怎样的，它们都完成了相同的功能：当细胞中葡萄糖供给充足时，它关闭相应代谢途径。再者无论是正调节还是负调节，协同调

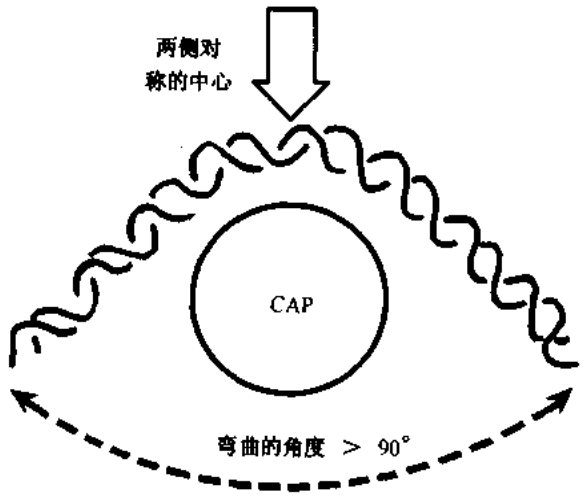


图 16－22 CAP 使 DNA 围绕对称轴弯曲 90°以上。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－27)

节通过调节蛋白结合位点的重复，可以扩展其分布的位点。

第二节　操纵子的其他调控形式

一、半乳糖操纵子

半乳糖操纵子（*gal* operon）是原核生物转录调控的一个典型。

1. 半乳糖操纵子的结构和功能

半乳糖也是 *E.coli* 的一种碳原，它的分解要涉及三种酶的催化：半乳糖激酶（galactokinase，K），半乳糖转移酶（galactose transferase，T）和半乳糖表面异构酶（glactose epimerase，E）。它们催化的反应如下：

$$Gal + Atp \xrightarrow{K} Gat-1-P + ADP + H^+$$

$$Gal-1-P + UDP-Glu \xrightarrow{T} UDP-Gal + Glu-1-P(UDP-Glu:尿苯二磷葡萄糖)$$

$$UDP-Gal \xrightarrow{E} UDP-G1$$

总的反应式为：

$$Gal + ATP \longrightarrow Glu-1-P + ADP + H^+$$

gal 操纵子结构如图 16－23、24 所示，其特点是：①有 2 个启动子：P_1 和 P_2，当有活性的 CAP 存在时 P_1 启动，其 －10 顺序位于 －12～－6，称为 $-10S_1$，转录的起始点为 +1。当 CAP 缺乏时 P_2 启动子启动，从 －5 开始转录，其 －10 顺序位于 －17～－11，称做 $-10S_2$；②gal 操纵子无 －35 顺序；③具有 2 个操纵基因 O_E 和 O_I，O_E 在上游，位于

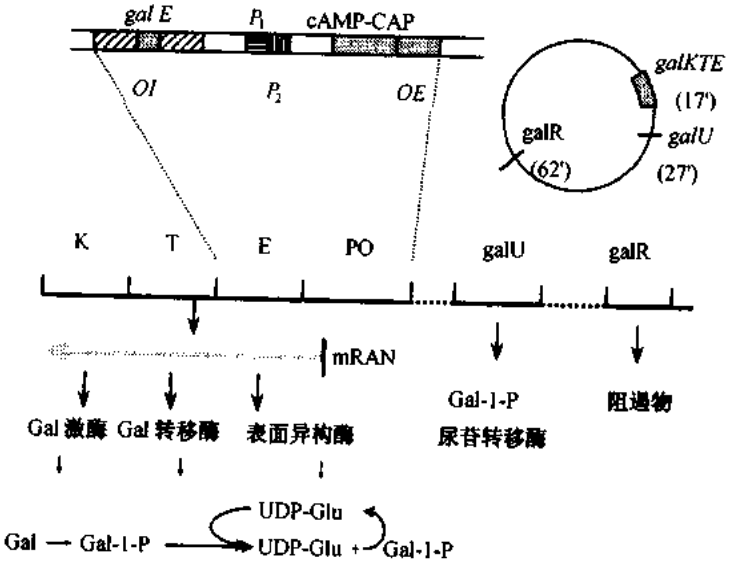


图 16－23　*E.coli* gal 操纵子的结构，在染色体上的分布及其产物催化的生化反应。

CAP 位点之内，O_I 在基因 *gal* 内部；无论是 O_E 还是 O_I 高启动子都有一段距离，不直接毗邻。

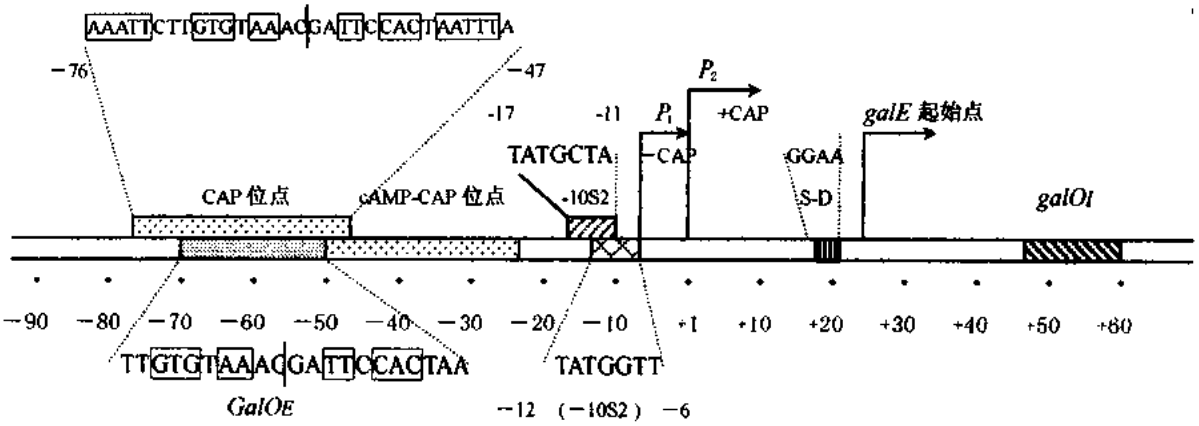


图 16－24　*Gal* 操纵子具有双启动子（P_1，P_2）和双操纵基因（*galOE*，*galOI*）

表 16－3　Glu 和 Gal 对 Gal 操纵子的调节

条件		表达
有 Glu	有 Gal	P_2 启动 S_2 开始转录 *gal E*，组成型表达
	无 Gal	O_E 和 O_I 相互作用，成环，转录只进行 20 碱基便停止
无 Glu	有 Gal	P_1 启动，3 个基因转录
	无 Gal	P_1 不启动

2. 调节

（1）如表 16－3 所示，当没有葡萄糖的条件下有 Gal 存在时，*gal* R（位于 62′）编码的阻遏物失活，CAP 结合在 －47～23 区域，RNA pol 结合在－ 10S_I 区，CAP 和 RNA pol 直接相互作用，使 P_1 顺利转录；当无 Gal 时 Gal R 结合在 Gal O_E 上，Gal O_E 具有回文顺序（TT GTG TAA AC I GATTCCACTAA）供 CAP 结合。它离启动子有一段距离，当 Gal R 结合在 *gal* O_E 上时不大可能像 *lac* 操纵子中 *lac* 阻遏蛋白那样去阻碍 RNA Pol 的结合，它阻断 S_I 的转录可能的两种方式：或是影响到 cAMP－CAP 和 RNA 聚合酶的作用而阻断；或通过干扰 cAMP－CAP 与 DNA 的相互作用来阻断。

在 Glu 存在的情况下，cAMP －CAP 含量少，当 Gal 存在，而无 Gal R 时，P_I 不能启动而 P_2 启动，RNA 聚合酶结合－10 S_2 顺序上，－10 S_2（TATGCTA）和－10S_1（TATGGTT）顺序相似，从 S_2 开始转录 *gal E* 而不转录另外的 2 个基因 *gal T* 和 *gal K*。这是由于 Gal 既可以作为碳源，同时 UDP－Gal 又是合成细胞壁的重要前体，因此无论 Glu 是否存在，只要有 Gal，*gal E* 总可以得到转录。

在 Glu 存在的情况下，若无 Gal 存在，Gal R 可结合在 *gal* O_E 和 O_I 上，并相互作用形成环。S_2 位点阻遏作用和 S_I 的阻遏机制完全不同，在上述条件下，即使有 Gal R 存在，P_2 仍不受阻遏开始转录（组成型），但由于 O_I 上也有 Gal R 的结合并且成环，故转录只进行 20 碱基便停止，这个过程如何发生尚不清楚。

（2）cAMP－CAP 的存在对 P_I 可以激活是正调控，而对 P_2 却是抑制，是负调控，其机制还没有搞清楚。

(3) Gal R 阻遏物对 P_1 和 P_2 两个启动子都是负调节，但在 cAMP－CAP 含量少时 P_2 仍可转录，其机制也不清楚。

(4) P_1 启动子与 RNA pol 的亲和力远远大于 P_2 启动子与 RNA pol 的亲和力。这可以解释为什么在没有 Glu，而有 Gal 存在时，P_1 转录，而 P_2 不转录。可能二者要竞争 RNA pol，而 RNA pol 在细胞中数量是一定的，由于 P_1 的亲和力大大超过 P_2，所以 RNA pol 几乎都结合到 P_1 启动子上，P_1 由于得不到 RNA pol 故不能启动。

二、阿拉伯糖操纵子

阿拉伯糖操纵子（ara operon）是指令合成阿拉伯糖分解代谢所需酶系的操纵子，它具有正、负调节的功能。

1. 结构和功能

阿拉伯糖的代谢是由 *araB*、*araA* 和 *araD* 基因所编码的三种酶的催化的（图 16－25）。其特点是：①AraC 蛋白是双功能的，单纯的 C 蛋白结于 $araO_1$（－100～－144），起到阻遏的作用；当 C 蛋白和诱导物 Ara 结合形成的复合体是 C^{ind}，，即诱导型的 C 蛋白，它结合于 *araI* 区（－40～－78）使 RNA Pol 结合于 P_{BAD} 位点（＋140），转录 B、A、D 三个基因；②C 蛋白结合 $araO_1$ 时也反馈性地阻遏了其本身的表达；③C 蛋白的两种状态（C^{ind} 和 C^{rep}）功能不同，结合的位点也不同。C^{ind} 结合于 *araI*；C^{rep} 可结合于 $araO_1$ 和 $araO_2$；④*ara* 操纵子的 C 蛋白还可以调节分散的基因 *araE* 和 *F*，因此此转录单位也称**调节子**（regulon）；⑤本操纵子有两个启动子 P_C 和 P_{BAD}，可以双向转录；⑥P_C 启动子和 $araO_1$ 重叠。

2. 调控

当 Glu 和 Ara 都存在时，C 本底转录，产生少量的 C 蛋白，结合于 $araO_1$（－106～－144），

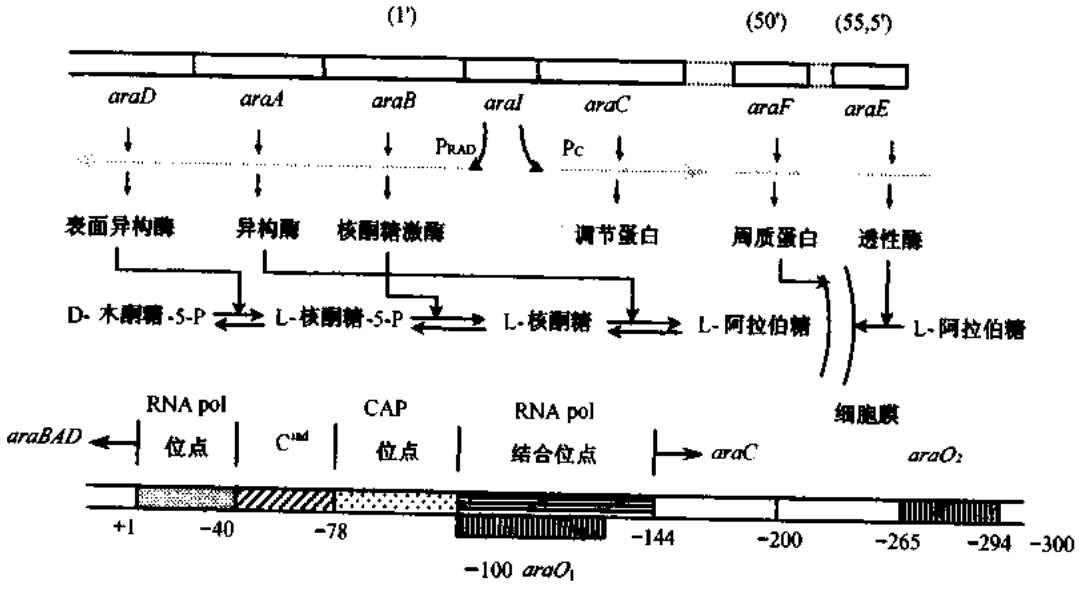


图 16－25　*ara* 操纵子的结构及其编码的酶所催化的生化反应。

使 RNA 聚酶不能结合 $araP_C$，使 *araC* 的转录受到阻遏(表 16－4)。

当有 Ara 存在，而没有 Glu 时，Ara 可作为糖源。此时 Ara 和少量的 C 蛋白结合形成了诱导型的 C 蛋白—C^{ind}，它作为正调控因子结合于 *araI*，促进了 *araPBAD* 的转录，产生了 3 种酶，促使 Ara 分解；

当 Ara 不存在或者用过完了，过量的 C 蛋白可以结合则 $araO_1$ 上，阻碍 RNA 聚合

表 16－4 *ara* 操纵子的调控

	条件	表达
有 Ara	有 Glu，无 CAP	araC 本底转录→少量 C^{rep}结合 O_1→阻止转录
	无 Glu，有 CAP	Ara + C→C^{ind}→*araI*→$araP_{BAD}$的转录
无 Ara	C^{rep}	C^{rep}过量→结合 O_1 阻遏 $araP_C$ 或结合 O_1 和 O_2 成环，阻遏 P_C 或 P_{BAD}

酶在此区域结合，从而关闭了操纵子；或者结合到 *araI*(－40～－78)和 $araO_2$ 上，彼此相互作用形成了环，阻遏了 P_{BAD} 和 P_C 的启动(图 16－26)。

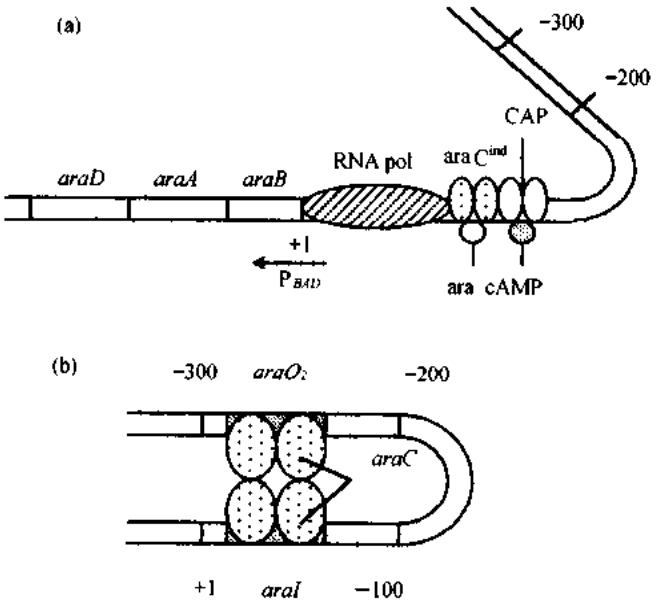


图 16－26 *ara* 操纵子的诱导(a)与阻遏。

三、色氨酸操纵子和衰减作用

乳糖操纵子是一个诱导系统，它控制的是分解代谢，底物小分子的存在诱导操纵子打开，合成一系列酶系，催化乳糖的分解。色氨酸操纵子正相反，它控制的是合成代谢，最终的产物是色氨酸，不难想像，只有在培养基中缺乏 Trp 时操纵子才能打开，而加入 Trp时将促进 *trp* 操纵子的关闭，也就是最终产物色氨酸或某种物质对转录将起到阻遏的作用，而不是诱导的作用。正由于 *trp* 操纵子是一个合成体系，不是糖的分解，因此和葡萄糖存在与否没有什么关系，在其操纵子中也就不存在 cAMP－CAP 位点。

色氨酸操纵子是合成代谢途径中最具有代表性的例子，Jacob 和 Monod 提出：在

E. coli 中由 5 个 *trp* 顺反子构成的基因簇形成了另一类型的操纵子，它不同于 *lac* 操纵子，色氨酸阻遏物只有和色氨酸结合才能具有活性，结合到操纵基因上，阻遏转录，这种类型控制途径也是在酶活性的水平来进行调节的。这种调节叫做**反馈抑制**(feedback inhibition)。在此代谢途径中的第一个酶(邻氨基苯甲酸合成酶)是由 *trpE* 和 *trpD* 基因编码的，它受到色氨酸本身的抑制。

1. trp 操纵子的结构和功能

trp 操纵子(图 16－28 所示)由 5 个结构基因 *trpE*、*trpD*、*trpC*、*trpB* 和 *trpA* 组成，在

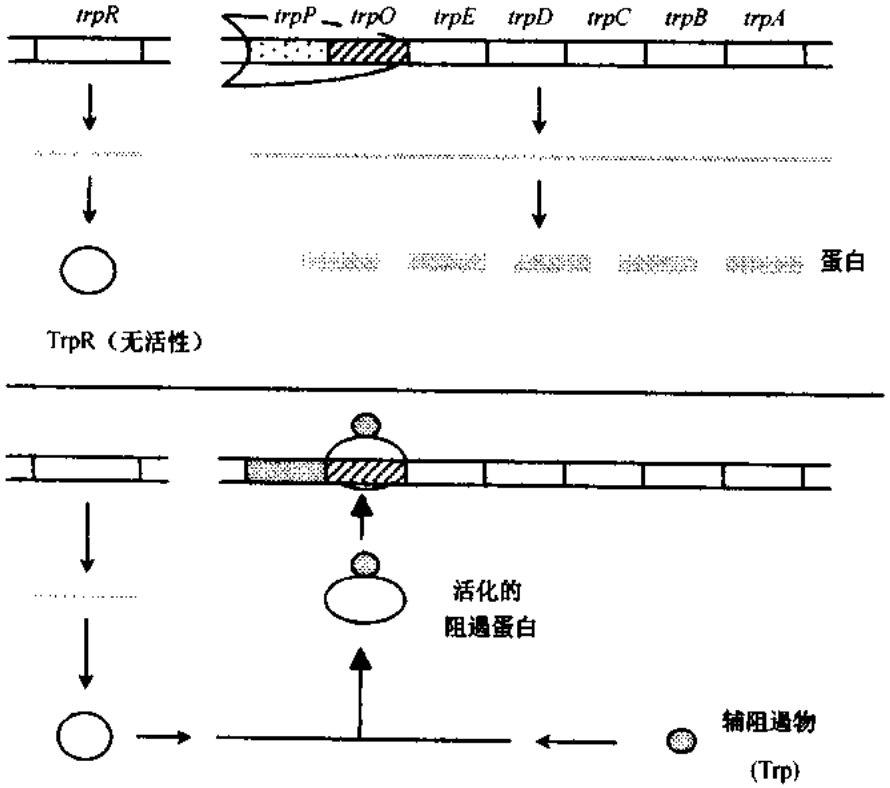


图 16－27　TrpR 被 Trp 激活后可阻遏 *trp* 操纵子的转录。(仿 B. Lewin: Genes Ⅳ, 1990, Fig. 13－16)

5′端是启动子、操纵子、前导顺序(*trpL*)和衰减子(attenuator)区域(*trpa*)。

其特点是：①阻遏物基因 *trpR*(89′)和结构基因(*trpEDCBA*)不紧密连锁；②操纵基因在启动子区域内；③启动子，操纵基因不直接和结构基因毗邻，而和前导顺序直接相联。④有衰减子结构，在合成代谢的操纵子的前导区内，存在着类似终止子结构的一段 DNA 顺序，称为衰减子，该顺序可以辅助阻遏作用进行转录调控，1981 年 yanofsky 提出了衰减模型。

2. 调节

和乳糖操纵子相同，无论是编码阻遏物的基因 *trpR* 发生突变，还是操纵基因发生突变，*trp* 操纵子将出现组成型表达。TrpR 是不能单独与操纵基因结合的，只有与色氨

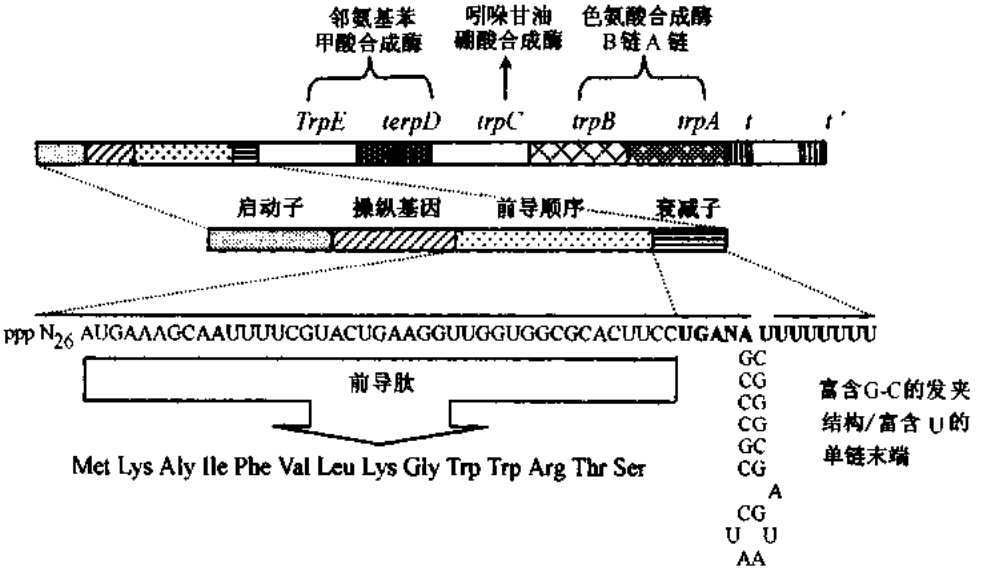


图 16－28　*trp* 操纵子含有 5 个结构基因和 1 个控制区。控制区由启动子、操纵基因、前导顺序和衰减子构成。前导区编码 14 个氨基酸，其中有 2 个是色氨酸。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－38)

酸结合后，构象发生改变，成为活性的阻遏物与操纵基因结合，使 RNA Pol 不能和启动子结合而抑制了转录。色氨酸就称为**辅阻遏物**(corepressor)。

Trp 操纵子的阻遏能力较低，仅是 *lacI* 产物的 1/ 1000，因此 *trp* 操纵还必须依赖别的途径来进行调节，以免在已有一定浓度的 Trp 时，还继续合成 Trp，这种途径正是下面要介绍的衰减作用。

第三节　转录的终止调控

Yanofsky 于 1978 年分离了多顺反子 *trp* 操纵子的 mRNA，测序后发现在第一个基因(*trpE*)的前面 5′端有一个长 160 碱基的序列，称为**前导序列**(leader sequence)。当此序列发生缺失突变时(缺失了 130～160 片段)产生的 mRNA 总是最高水平的，yanofsky 称这个元件为**衰减子**(attenuator)(图 16－29)。当 Trp 存在时，由于衰减子的存在导致了 mRNA 转录速率速率下降。但前导序列中 1～130 的这段序列的作用是什么呢？Yanofsky 提出了一个假设解决了这个问题。

衰减的共同特点是某些外部因素控制着衰减子的发夹形成，它可使转录终止。如发夹形成终止子，它将阻止 RNA 聚合酶转录结构基因，若不形成发夹结构，RNA 聚合可以通读终止子，结构基因可以得到表达，转录开始于基因簇的左端，邻接的是操纵基因，*trpR* 就结合于此区。在操纵基因和编码区之间是前导序列。结构基因的转录部分是终止于内部的终止子 *trpt*，它在编码区外，长 36bp，后面间隔约 250bp 还有一个 ρ－依赖性终止子 *trpt*。现已证明了这个终止位点对于 *trp* 操纵子来说是不需要的。无论在 *E. coli* 还是在沙门氏菌中这种操纵子的本质都是相同的。

(1) 前导序列可能编码了 14 个前导肽。

(2) 衰减子(内部终止子)为结构基因的转录设了一道关卡，无论在体内还是在体外，RNA 聚合酶在此处终止都只产生 140nt 的转录本。

衰减子上的终止可对细胞内 Trp 的水平作出反应(图 16－29)。当存在适量的 Trp

时，转录终止，而在缺乏 Trp 时，DNA 聚合酶可持续到达结构基因。

阻遏与衰减以相似的方式对 Trp 存在的水平作出反应，当 Trp 存在时，操纵基因被阻遏；但由于阻遏的能力不强，使大部分的 RNA 聚合酶得以逃脱，继续向前，可是最终

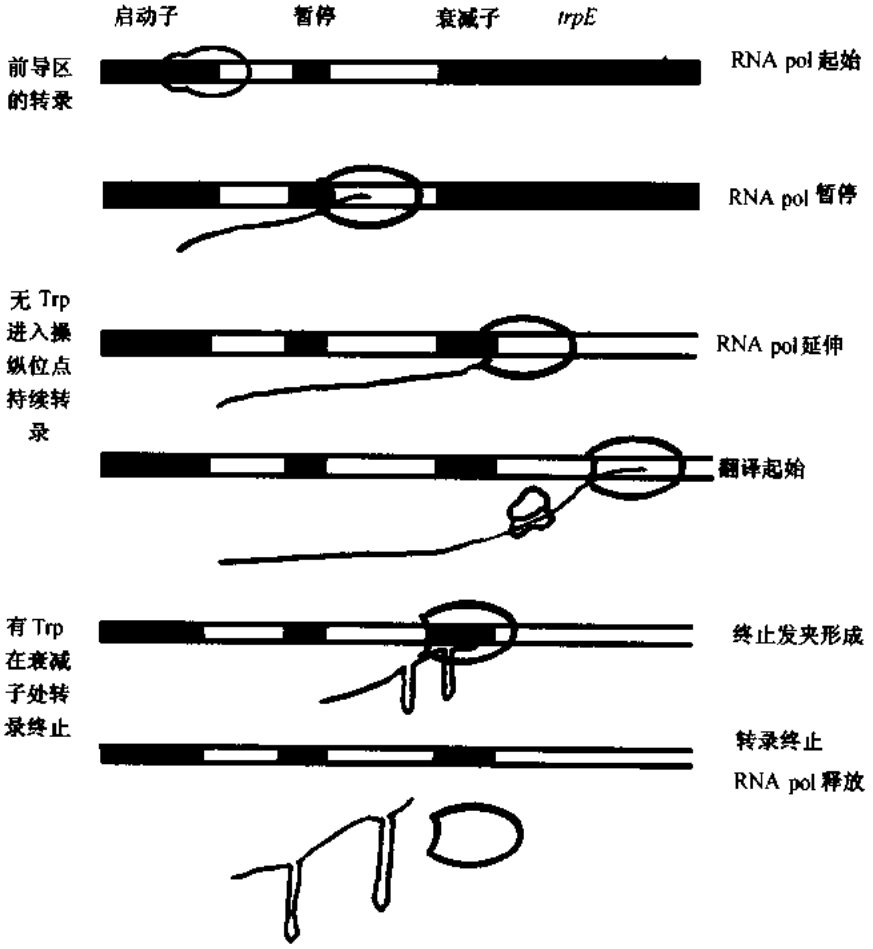


图 16－29　衰减子对转录的调节。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－39)

还是在衰减子处被拦了下来。当去除 Trp 时，RNA pol 能自由地通过启动子，也不大会过早被终止。

衰减时转录的作用增加 10 倍左右：当 Trp 存在时终止是有效的，而衰减子仅允许约 10%的 RNA 聚合酶前进。在缺乏 Trp 时，衰减子允许所有的聚合酶通过。由于阻遏蛋白的释放，在转录起始增加约 70 倍，这样二者的作用使操纵子的调节范围在 700 倍左右。

衰减子如何能终止转录对 Trp 的水平作出反应呢？前导肽的序列说明了这个机制。图 16－28 表明前导序列编码 13 个密码子。这个序列是否翻译成导肽呢？虽然在体内测不出产物，但可能由于其不稳定所致，我们已知道其核糖体结合位点是有功能的。

导肽含有 2 个 Trp 残基，在 *E. coli* 蛋白中 Trp 是属于稀有氨基酸，当细胞中 Trp 用完时，核糖体开始翻译导肽，但他们达到 Trp 密码子时就停下来。mRNA 的序列表明核糖体的延宕(ribosome stalling)影响了在衰减子处的终止。

前导序列能形成不同氨基酸配对的结构，核糖体控制这些结构的转变而穿越过前导区，这些结构决定了 mRNA 是否能提供具有终止功能的二级发夹。

图 16－30(a)表示这些结构，首先可以看出 1 区和 2 区可以碱基配对；3 区和 4 区

也可以碱基配对。3区和4区配对产生的发夹，其末端存在8个U的寡聚U序列，这实质上是内部终止信号，RNA产生的这种结构可能取决任何外部的影响。如果1区不和

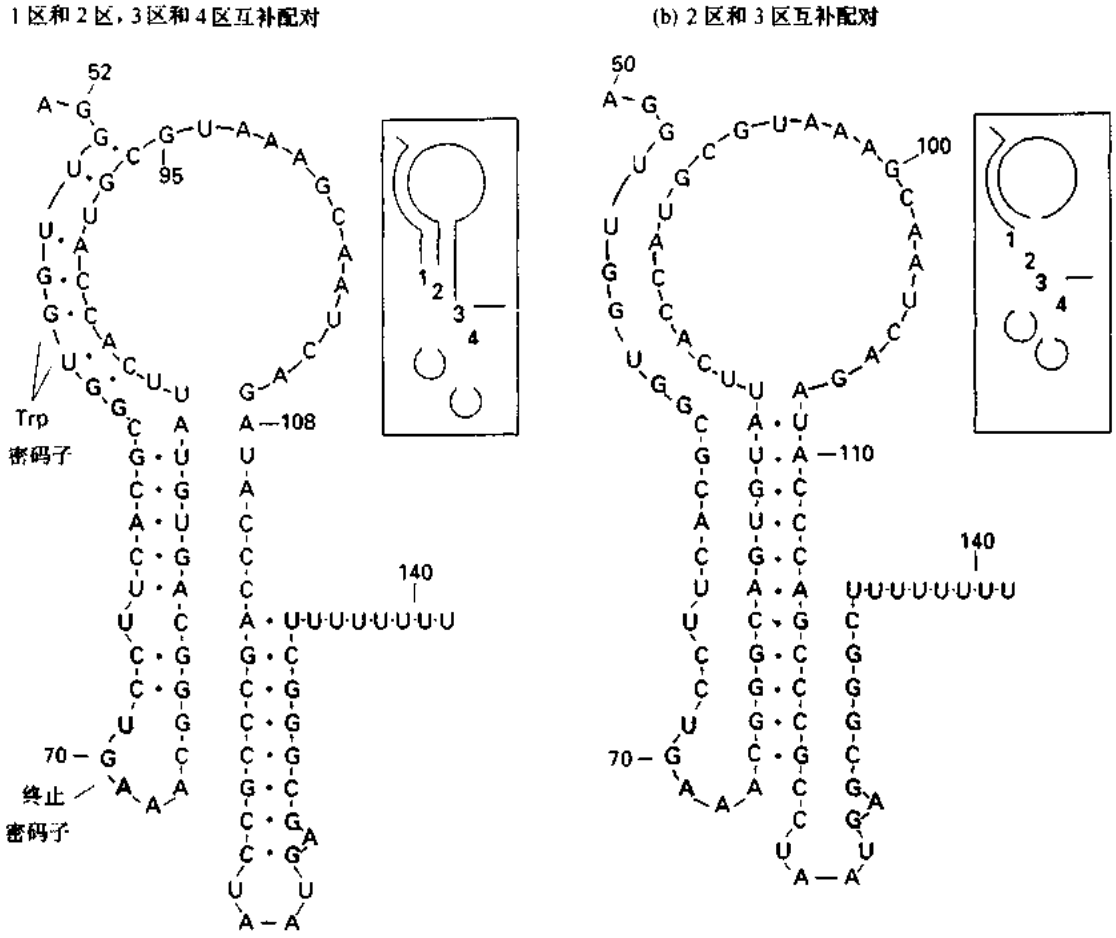


图16－30 *trp* 前导区能产生不同的碱基配对构象。它有4个碱基配对区。1区和2区，2区和3区，3区和4区都可互补配对。(a)表示1区和2区，3区和4区互补配对；(b)表示2区和3区互补配对，使1区和4区呈不配对的构象。(仿B. Lewin: Genes Ⅵ, 1997, Fig. 16－19)

2区配对就会形成不同的结构(图16－30b)。在这种情况下，游离的2区可以和3区配时，那么4区就失去了配对伙伴，而保持单链状态，这样就不能形成终止子的结构。

图16－31表明核糖体所处的位置能决定不同结构的形成。衰减作用以这种方式仅仅在Trp充足的时候来终止转录。这一重要的性质是由在前导序列中Trp密码子的位置来确定的。当Trp存在时，核糖体能合成前导肽，核糖可以沿着前导序列的mRNA向前延伸，直到终止密码子UGA才停下，此UGA位于1区和2区之间，核糖体覆盖了1区和2区的部分区域，2区余下的部分只能和3区的少数碱基配时，3区余下的部分仍可以和4区配对形成发夹的终止子结构(图16－31b) 可以终止转录。在此情况下，RNA pol在衰减子上终止 。当缺乏Trp时，核糖体进行到1区的Trp密码子时出现了图16－31a的情况，1区完全被核糖体所占据，无法与2区配对，这样在4区被转录前2区就和3区配对，使 整个的4区保持单链状态。而不能形成终发夹，RNA pol就能持续转录而越过。

通过衰减子控制转录的终止仅发生在原核生物代谢基因。核糖体暂停的位置决

定了二级结构形成的方式，从而控制转录的终止，因此必需转录和翻译偶联才行。当 RNA pol 达到终止位点时，前导肽的翻译必须同时发生。控制时间的关键是存在一个位点导致 RNA pol 停留在前导序列的第 90 个碱基处。RNA pol 一直停着，直到核糖体

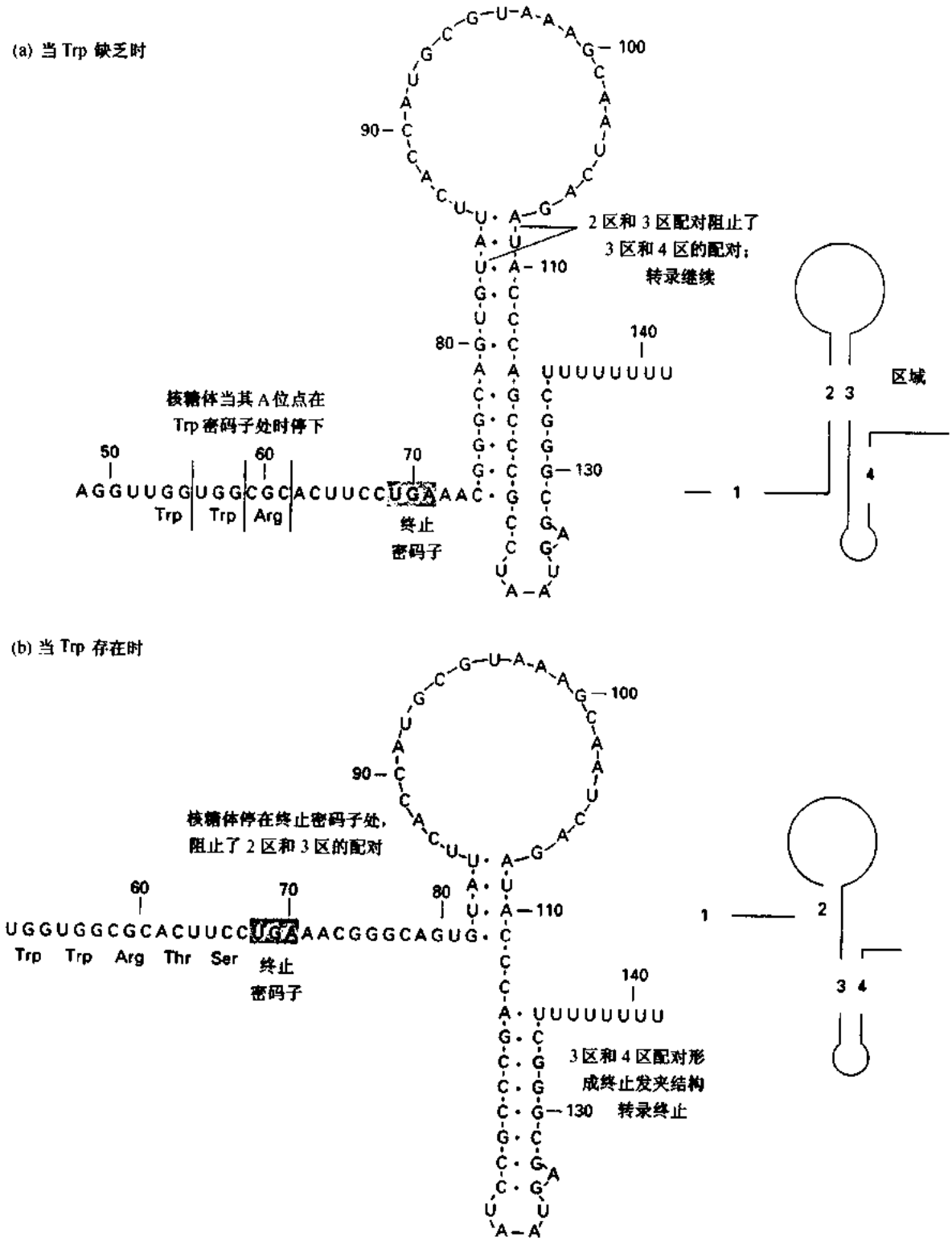


图 16－31　RNA 聚合酶在衰减子处的衰减依赖于核糖体在衰减子上的定位，此位点取决于是否能使 3 区和 4 区互补配对，末端形成发夹结构。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 16－18）

翻译了前导肽。然后多聚酶被释放并在衰减位点解离下来。当它到达时，衰减区域的二级结构已经确定了。

这个机制意味着 Trp－tRNA 的供应不足，衰减对细胞中 Trp 的需求作出直接的反应。

衰减作为细菌操纵子中的一种调控机制其广度如何呢？它至少存在于氨基酸合成代谢的 6 种操纵子中。由氨基酸水平来反馈调节蛋白质合成，这对经济合理地合成酶可能是有利的。

其他一些氨基酸合成操纵子也有衰减机制。它们的分子信息表明在各种情况下，衰减的机制是相似的。十分有趣的是某些操纵子仅使用衰减而不用阻遏，例如没有发现组氨酸的阻遏蛋白。

第四节　RNA 聚合酶转录起始

一、σ 亚基的替换

在枯草杆菌（*B. subtilis*）中 σ 因子广泛地用于转录起始的调节，现知道有 10 种不同的因子。有的存在营养期细胞中，即在被噬菌体感染的时期，或者从营养生长转变成孢子形成期。

在处于正常营养生长期的枯草杆菌中发现的 RNA 聚合酶与 *E. coli* 的 $\alpha_2\beta\beta'$ σ 的结构相似，已知 σ 因子的分子量为 43kDa，因而以 σ^{43} 或 σ^A 来表示。它所识别的启动子带有的保守序列，与 *E. coli* σ^{70} 识别的相似。各种不同的聚合酶含有不同的 σ 因子，但是数量很少。各种聚合酶识别不同启动子的 －35 和 －10 序列。

从一套基因的转录转换为另一套基因的表达是噬菌体感染的共同特点。噬菌体的发育涉及到感染周期的改变。这些改变通过噬菌体编码的 RNA pol 的合成来完成，即通过噬菌体编码的控制细菌 RNA 聚合酶的附属因子（包括新的 σ 种类）来完成。在枯草杆菌被噬菌体 SPO1 感染中通过产生新的 σ 因子来控制的。SPO1 的感染周期通过基因表达的三个阶段。在感染的瞬间，噬菌体的早期基因被转录了。在 4～5 分钟后早期转录停了下来，中期基因又开始转录。再过 8～12 分钟中期基因的转录被晚期基因所取代。

早期基因被宿主菌的全酶所转录。宿主基因启动子也能被 RNA 聚合酶 $\alpha_2\beta\beta'\sigma^{43}$ 所识别。

噬菌体基因的表达对于中期和晚期基因的转录是必要的。有 3 个调节基因叫 *28*，*33* 和 *34*，它们控制要转录的程序。它们的功能如图 16－32 所示。调节的方式是一种级联调控。在级联调控中宿主的酶转录早期基因，这些基因的产物是转录中期基因所必须的。而两个中期基因编码的产物又是晚期转录所必须的。

早期基因 *28* 的突变体不能转录中期基因，基因 *28* 的产物（称为 gp28）是分子量 26kDa 的蛋白，它取代核心酶上的 σ 因子。

这种替代对从早期基因的表达转为中期基因表达是必要的。它形成的全酶不再能转录宿主的基因，而能特异转录中期的基因。现在还不清楚 gp28 怎样取代 σ^{43}。

两个中期基因涉及到一下步的转录。无论是基因 *33* 还是 *34* 若发生突变将会阻止晚期基因的转录。这些基因的产物是 13kDa 和 24kDa 的蛋白，它们取代了核心的酶

上的 gp28，现在也还不知道 gp33 和 gp34 怎样排除 gp28 的（或者排除任何残余的宿主 σ^{43}），但它们一旦结合到核心酶上，它们就只能在晚期启动子上起始转录。

σ 因子的相继的取代具有双重的后果，每次亚基的改变使 RNA 聚合酶能够识别一组新的基因，而不再识别先前的基因。由于 σ 因子的转换使 RNA 聚合酶的活性发生

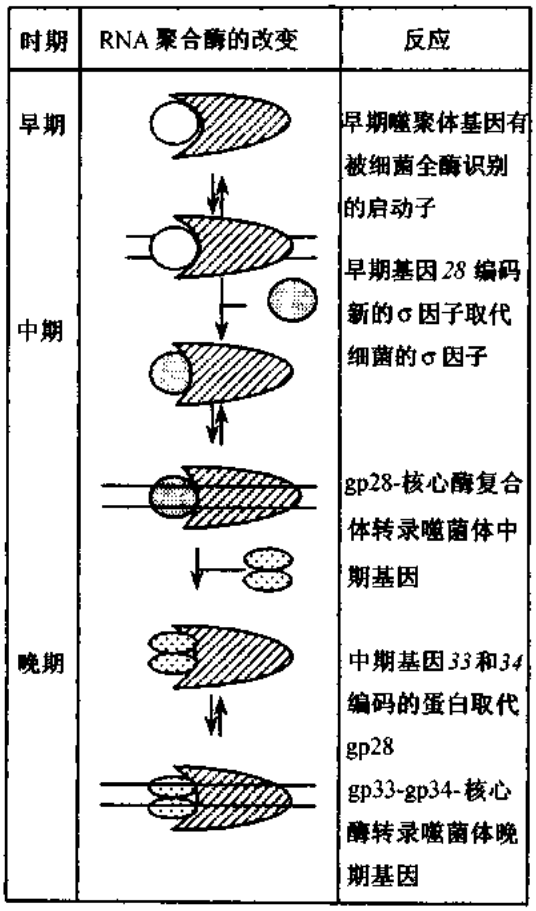


图 16-32　SPO1 噬菌体的转录由各种 σ 因子来控制。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11-22）

了改变。可能所有的核心酶都是短暂地和不同的 σ 因子结合，这种变化的程序是不可逆的。

几乎大部分 σ 因子转换的例子都发生在**孢子形成**（sporulation）中。不同生活途径的选择对某些微生物来说是有利的，在**营养期**（vegetative phase）来说，对数生长可导致培养基中营养物质的耗尽。这引发了孢子形成（图 16-33）；它包括以下几个阶段：①DNA 复制；②在细胞的一端基因组被分离；③分离的基因组外包上一层孢子外壳。在孢子形成的第二阶段，细胞产生了两个独立的被分隔的部分，这就是母细胞和前孢子（forespore）。这个过程约花 8 小时，它可以被看成为是一种原始的分化类型。在这种类型中，现代细胞产生两种发育命运不同的子细胞，一个是母细胞，一个是前孢子，当前孢子被释放时，母细胞最终被裂解，孢子的结构完全不同于原来的细菌。

孢子的形成涉及到细菌生物合成活性的剧烈的变化。这些变化和很多基因有关。基本的调控仍在转录水平。在孢子形成时，一些营养期行使功能的基因被关闭了，但大部分仍继续表达。另外孢子形成的特异基因仅在这一阶段表达。在孢子形成结束时

约有 40% 的细菌 mRNA 对孢子形成是特异的。

孢子形成的 RNA 聚合酶在孢子形成的细胞中成为活性状态，它们含有的核心酶和营养细胞中的相同。但附属蛋白却不同于营养细胞的 σ^{43}。图 16－34 表明了这种转

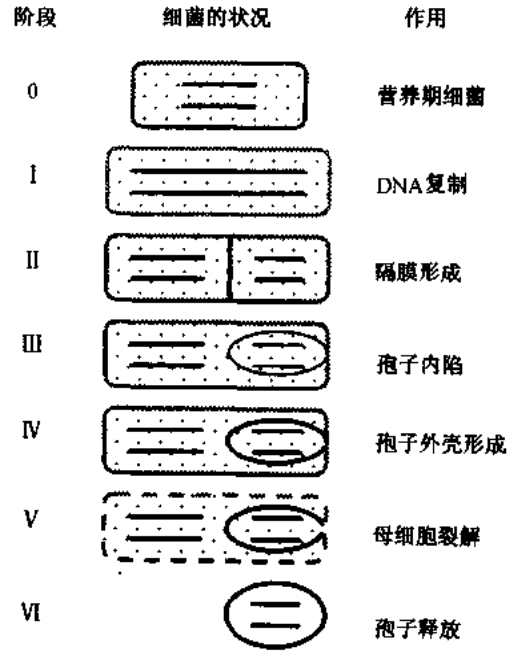


图 16－33 细菌孢子的形成和释放。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11－23)

录的特异性改变。其原理是在每一间隔中存在着 σ 因子连续地被新的因子所决代，导致了不同组基因的转录。为了协调前孢子和母细胞中转录的时间，必须沟通这两个部分。

当外界环境条件能触发"磷酸化传递"时，孢子形成的级联调节就起始了。在磷酸化传递中磷酸化沿着各种蛋白传递，直至到达 SpoOA(多种基因的产物涉及此过程，这种复杂性可能反应在触发孢子形成中需要避免发生错误)。 SpoOA 是一种转录调节物，其活性受磷酸化的作用。在磷酸化状态时，它激活两个操纵子转录。而每个操纵子的转录是由不同于宿主的 RNA 聚合酶催化的。在磷酸化 SpoOA 的指导下，宿主酶利用了一般的 σ^{43} 来转录编码 σ^{F} 的基因。而宿主酶在较小的因子 σ^{H} 的直接指导下转录编码 σ^{E} 的基因。在中隔形成前，这两种新的 σ 因子产生了，但到晚些时候它们才被活化。

σ^{F} 只有在前孢子的间隔部分才有活性，而在母细胞中它的作用却被抑制了。在孢子形成的开始，在前孢子中 σ^{45} 被 σ^{F} 取代了。在 σ^{F} 的指导下，RNA 聚合酶转录第一套取代了营养期基因的孢子形成基因。营养期基因是先前转录的。由于 σ^{F} 产生量很少，因此某此营养期的酶仍保留存在于孢子形成时。被取代的 σ^{45} 并未破坏，从孢子形成的细胞中的抽提液中 σ^{45} 仍可得到恢复。通过两种调节事件才使 σ^{F} 激活。在前孢子中有一种 σ 因子：σ^{G}，它是早期孢子形成基因的产物，它使 RNA 聚合酶在前孢子中转录晚期基因。另一种早期孢子形成基因的产物只负责与母细胞间隔进行沟通。一种信号(通过间隔的蛋白膜)的传递来激活 σ^{E}，σ^{E} 在先前就已以前体的形式(pro－σ^{E})被分成，进一步被剪切后才有活性。

当 σ^{E} 依次导致 σ^{K} 基因转录时，上述级联调控仍在继续。σ^{K} 活性的产生是十分复杂的，首先它的基因需要通过重组才能产生。这个因子也是作为一种无活性的前体蛋白(pro－σ^{K})被合成的。它被一种蛋白酶所激活，一旦 σ^{K} 有了活性，它就取代 σ^{E} 并导致

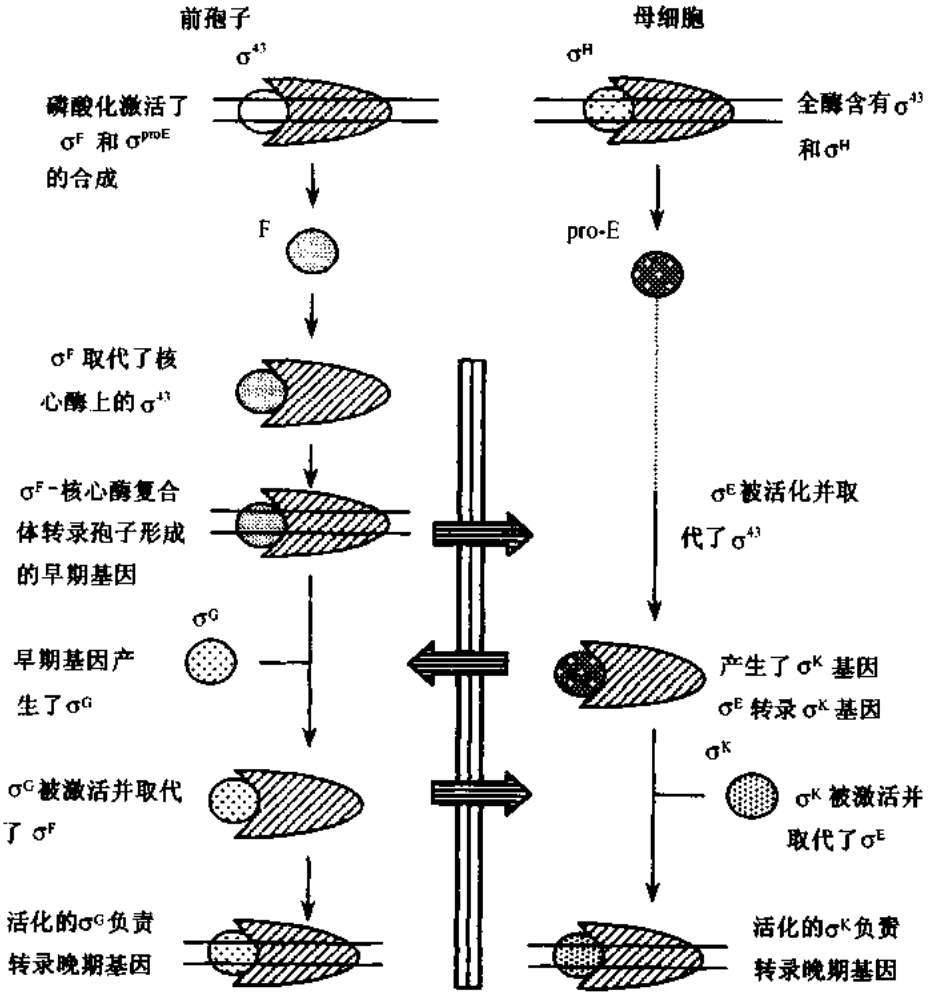


图 16-34　孢子形成涉及到 σ 因子的改变，从而控制 RNA Pol 起始转录的特异性。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11-24)

了母细胞中晚期基因的转录。这些事件在母细胞和前孢子两部分的定时是由一些信号协调的。在前孢子中 σ^{G} 的激活是依赖于母细胞中发生的一些事件。依次激活 σ^{G} 可以产生一种信号，这种信号穿越过间隔去激活 σ^{K}。

孢子形成由两种级联调控控制，在级联调控中每一间隔部分的 σ 都相继被激活，每个 σ 因子指导一套特殊的基因合成蛋白质。这两种级联调控通过一种信号从一个间隔部分传递到另一个间隔部分来彼此沟通协调。当一种新的 σ 因子被激活，原来的 σ 因子被取代，这样通过转移 σ 因子来打开基因或关闭基因。各种因子结合的 RNA 聚合酶指挥一套靶基因表达。这些因子的量有效地影响基因表达的水平。在任何时候都有一个以上的 σ 因子被激活。某些 σ 因子的特异性是重叠的。现在还不清楚为什么每个一个 σ 因子能替换前一个 σ 因子。

二、σ 因子的替代可以控制起始

E. coli RNA 聚合酶转录全部细菌的基因，它能广泛识别各种启动子。但与不同启

动子相连结的基因是以不同的水平和在不同的场合被转录的。在特殊启动子上的启动转录的能力是由很多的附属因子的帮助或者和酶的相互作用来控制的。几乎没有一个例子是核心酶亚基本身发生改变来进行调节的。这种机制并非有一种"偏爱",而是由于这种改变要使 RNA 聚合酶能特异识别一种类型的启动子，并阻止其他类型的启动子的转录,将减少聚合酶的机动性。在核心酶的亚基只负责链的延伸,而由 σ 因子来负责位点的选择，每种不同类型的启动子都有自己的特异性，那么是否也需要多种不同类型的 σ 因子呢?

在有些情况下的确存在 σ 因子的变换,例如细菌中在孢子形成时就需要变换 σ 因子。*E.coli* 虽无需经历如此戏剧性的变化，但它进入静止生长期时也要经过相当复杂的代谢变化。在一个很长的时间内 *E.coli* 只有一个 σ 因子——σ^7，现在又有了新的发现,如表 16－5 所示,这些 σ 因子的命名或根据其产物的分子量,或根据基因的名称。

表 16－5　*E.coli* σ 因子识别不同保守顺序的启动子

基因	分子量	功能	－35 顺序	间隔(bp)	－10 顺序
rpoD	70KD	普遍	TTGACA	16～18	TATAAT
rpoH	32KD	热休克	CCCTTGAA	13～15	CCCGATNT
rpoE	24KD	热休克	?	?	?
rpoN	54KD	氮饥饿	CTGGNA	6	TTGCA
fliA	28KD	产生鞭毛	CTAAA	15	GCCGATAA

B. Lewin: Genes Ⅵ, 1977, Table 11.1

σ^{32}(也称 σ^{N})和 σ^{54}(或称 σ^{H})轮流交替被激活,对环境的改变作出反应。σ^{28}(或称 σ^{F})用于正常生长情况下鞭毛基因的表达。但环境发生改变时其表达水平会发生改变。

当环境温度增加时,*E.coli* 改变了它们的转录模式以对"**热休克**(heat shock)"作出反应。在很多生物中,无论是原核还是真核生物,对热休克反应是一种共同类型。温度升高,通用蛋白质合成开始关闭或下降,而新的一类蛋白却开始合成。这些新的蛋白是**热激基因**(heat shock genes)的产物,它的作用是保护细胞,对抗环境的变化。它们的合成是对新的条件如热休克作出反应。各种热休克蛋白都属于**分子伴侣**(chaperones)。在 *E.coli* 中，17 种热休克蛋白的表达都是通过转录的改变触发的，*rpoH* 基因是调节转录热休克反应所必需的。其产物是 σ^{32},其功能是作为一种替换的 σ 因子,可开启热休克基因的转录。

另一些 σ 因子是用于氮饥饿时。*E.coli* 细胞中含有一些小分子蛋白，现称为 σ^{54}，当培养基中缺乏氮时就要用到它。在这种条件下基因打开,允许利用氮源。在细菌中广泛地发现这种 σ 因子的类似物。

另一种环境引起 σ 因子变换的情况是 σ^{F}，它在细胞中存在的量很少，但可导致 RNA 聚合酶转录那些和趋化性有关的基因以及鞭毛的结构基因。在枯草杆菌中它的类似物是 σ^{D},它可控制鞭毛形成及游动基因。带有相同启动子特异性的因子存在于很多特殊的细菌中。

多种形式的 σ 因子的存在就进一步使人们思考 σ 对核心酶的作用是什么?什么在负责控制 σ 因子和核心酶的连结。怎样有效地决定全酶类型的调节?

我们现在完全不了解在生长的细菌中 σ 因子之间相互关系的调节机制。在对环境作出应激反应时(如热休克或氮饥饿)产生新的 σ 因子,并使用它。假设在热休克时基因转录的选择依赖于普通的 σ^{70} 和热休克 σ^{32} 之间的平衡，这两个 σ 因子共同竞争有

效的核心酶，σ^{32} 蛋白是不稳定的，它可以迅速地增加或减少，它的不稳定性成为调节的重要性能。

每种 σ 因子都能导致 RNA 聚合酶起始转录一套特殊的启动子。而每套启动子可以通过单一序列元件来加以鉴别。每种类型启动子的序列又是通过 RNA 聚合酶附加的 σ 因子直接识别的。在 *E. coli* 和枯草杆菌孢子形成中 σ 因子所负责的那些基因的识别可以推断起始识别的一般标准。

启动子对于每种酶来说有一个重要的特点，那就是相同的大小和与起始位点的距离。它们通常仅在 - 35 和 - 10 序列的中心有保守序列。这种保守序列无论是 - 35 还是 - 10 对于每套启动子来说是不同的。一种特殊的 σ 因子仅能识别特有的一套启动子，因此不同蛋白的转录都是特异的。这样一个 σ 因子被另一 σ 因子所取代，原来的一套基因就会被关闭，新的一套基因就会被打开。

不同的 - 35 和 - 10 保守序列被含有不同 σ 因子的全酶所识别。这就直接表明亚基本身必须直接接触此区的 DNA，只有被 σ 因子识别了才能通过核心酶的构象间接地发挥功能。有可能 σ 能导致核心酶产生系列构象中的一种，而每种构象对不同的启动子序列来说都是特异的。这表明了在 σ 因子和核心酶之间有一种共同性的关系，σ 因子和启动子的 - 35 及 - 10 序列产生关键性的接触，以这种方式来定位。但现在还没有解决 σ 作为一个小分子多肽怎么能覆盖 20bp 以上跨度的 DNA。

比较各种细菌 σ 因子区域的序列发现它们是保守的。我们预期这将有重要的意义。σ^{70} 有 4 个保守区，如图 16 - 35 所示，2 区和 4 区启动子的 - 10 及 - 35 序列相互作用，2 区的其他部分是 DNA 熔解所需要的。第 361 ~ 390 残基是和核心酶结合的区域。N - 端区是阻止 2、4 区在缺乏核心酶的情况下与 DNA 结合。

σ 因子和启动子 - 35 及 - 10 直接接触的证据来自 σ 保守区的突变。当启动子特殊位点发生突变也会阻止 RNA 聚合酶的识别。而在 σ 因子上发生相应的补偿突变时就允许聚合酶用于突变的启动子，从这个实验中不难推测出 DNA 中有关的碱基对和那些被取代的氨基酸接触。图中（16 - 35 ）表明 2 和 4 是涉及到和 - 10 及 - 35 接触的，这两个区域在蛋白质中以 α - 螺旋的形式存在。

σ 的其他部分没有发现单独的功能。2 区的某些部分具有能与单链核酸结合的蛋白相似的序列，推测有可能涉及解链。在 2 区之前和 2 区开始部分功能有些重叠，看来这部分是和核心酶相互作用的区域。

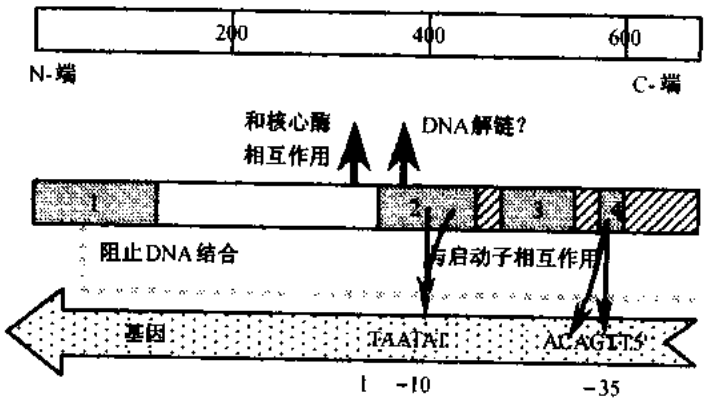


图 16 - 35　*E. coli* σ^{70} 因子鉴别出的 4 个保守区。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 11 - 20）

当 σ^{70} 的 N 端区域被切除时，这一段截短了的蛋白就能特异地与启动子序列结合，而完整的 σ^{70} 却不能，这表明在 σ^{70} 游离状态时 N 端区域可封闭 DNA－结合功能区，但 σ^{70} 和核心酶结合后改变了 σ^{70} 的构象，因此能与 DNA 结合。

在蛋白中用 α－螺旋模体来识别双链 DNA 序列是共同的特点。图 16－36 表明在一个 α 螺旋上相隔 3～4 个氨基酸的位点位于双螺旋的同一表面上，而且在某一位点与碱基相连接。在 *E.coli* 和枯草杆菌主要的 σ 因子中，其 Thr 和 Arg 在双螺旋的相关位点，并负责和－10 序列中的头两个位点相接触。还不清楚－10 序列的其余部分如何接

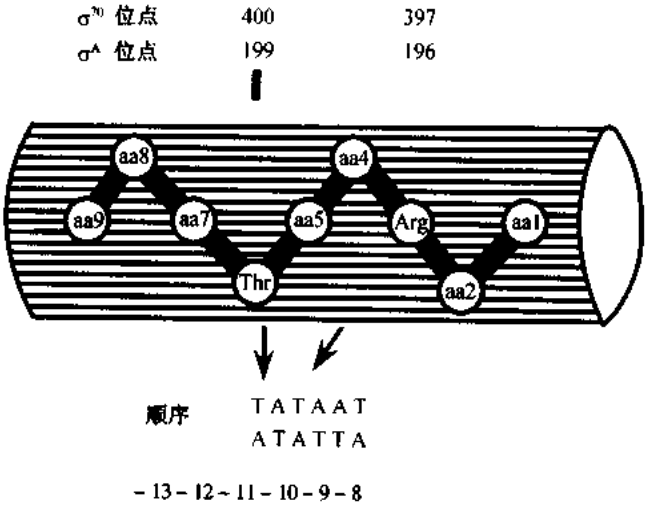


图 16－36　在 *E.coli* σ^{70} 或枯草杆菌的 σ^{A} 的 2.4 α－螺旋中相关位点的氨基酸可以和－10 顺序的前两个碱基接触。（仿 B.Lewin:Genes Ⅵ,1997, Fig.11－21）

触。实际上 σ 和 DNA 之间的接触范围以及核心酶和 DNA 之间的接触关系还有待于建立。

三、修饰核心酶、替换 σ 亚基

T4 噬菌体蛋白质合成的时序调节采用了与 *E.coli* 及枯草杆菌不同的策略，那就依赖本身携带的蛋白对宿主的核心酶加以修饰，改变其和 σ 亚基的结合能力，乘机将本身编码的蛋白取代原来的 σ 亚基，从而改变 RNA 聚合酶的识别能力。

T4 噬菌体在生活周期的不同阶段合成不同的 mRNA。其主要是早期 mRNA 和晚期的 mRNA。早期转录的有很多基因，它们编码有关 DNA 合成、RNA 的转录调控、重组等的酶；晚期转录的是一些编码噬菌体结构蛋白，粒子装配以及裂解宿主等的基因。

T4 感染伊始是利用宿主的 RNA 聚合酶转录第一批早期 mRNA；T4 的头部含有几种小分子蛋白，称为**内部蛋白**。当 T4 感染不久，这些蛋白伴随着 DNA 一道注入到细菌中，其中有一种转换蛋白是 **ADP－核糖转移酶**（ADPRT），它可将 ADP－核糖（ADPR）连接到宿主 RNA 聚合酶的 α 亚基上的精氨酸胍基上（图 16－37）使其被修饰，加入一个以上的 ADPR，从而降低了 RNA 核心酶与 σ 因子的结合能力，而增加了它与 T4 噬菌体编码的蛋白 Mot 的亲和力，这样 Mot 蛋白取代了 σ 亚基，使新的 RNA 聚合酶不再识别寄主的基因和 T4 早期基因的启动子。而能识别下一批基因的启动子。到转录晚期，RNA 聚合酶被进一步地修饰，其两个 α 亚基各结合了一个 ADPR 分子及 4 个修饰性蛋

白，有时称其为**超修饰**（hupermodified）的 RNA 聚合酶。即使如此它仍不能转录晚期基因，要等到 DNA 复制时才能转录，形成了一种程序化的控制。实际上直到复制达到一

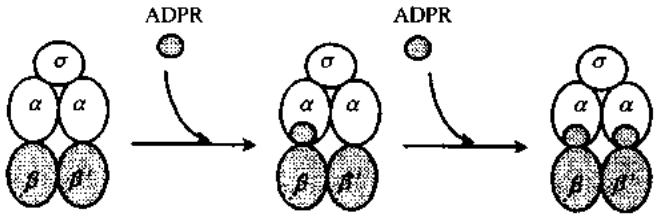


图 16－37　T4 的 ADPRT 修饰 *E. coli* 的 RNA 聚合酶。（仿 D. Freifelder: Molecular Biology, 1983, Fig. 15－16）

定阶段才需要晚期基因编码的产物。对于复制时晚期转录的调控机理还不清楚，可能是改变了晚期基因的构象，使其易于转录启动。

四、RNA 聚合酶的代换

T7 噬菌体在转录的时序调控中不是采用 σ 因子的级联取代，也不是像 T4 噬菌体那样对核心酶进行修饰，而是根据感染特点采用了 RNA 聚合酶的代换来进行时序转录的调控。

T7 是 *E. coli* 的烈性噬菌体，基因组长为 39,936nt，序列已知。整个基因组可能具有 55 个基因，其中 44 个基因的产物已被鉴别，30 种产物的功能已搞清楚。

从图 16－38 中可以看出转录可分为 3 组，30℃ 整个生活周期约为 25 分钟，但其 DNA 的注入很慢，整个的过程约 10 分钟，而其他的噬菌体仅 1 分钟即可完成。这也是一种调控的手段，因未注入宿主细胞中的 DNA 是不能转录的。在感染 6 分钟后，*E. coli* 的 RNA 合成体系全部被 T4 的合成体系所取代。

在感染后的前 6 分钟，T_7 噬菌体的 RNA pol 尚未表达，因此仍使用了宿主 *E. coli* 的 RNA 聚合酶，转录的基因主要有 10 个（*0.3, 0.4, 0.5, 0.6, 0.65, 0.7, 1, 1.1, 1.2, 1.3*），其 *0.3* 基因编码宿主限制酶的抑制蛋白，使 T7 进入宿主免遭降解；*0.7* 基因编码一种蛋白激酶，能将 *E. coli* 的 RNA 聚合酶磷酸化，为抑制宿主 RNA pol 作准备；基因 *1* 编码 T7 RNA pol 聚合酶，分子量为 98kDa，是一条单链肽，它所识别的启动子是由 23bp 组成的保守序列（－17～16）。

晚期转录分为二组（Ⅱ和Ⅲ）。感染后 6～12 分钟后 T7 就利用自己的 RNA 聚合酶转录第Ⅱ组转录物。这一组约含（*1.4, 1.5, 1.6, 1.7, 1.8, 2, 2.5, 2.8, 3, 3.5, 3.8, 4, 5, 5.7, 6*），其中对转录调控有关系的是基因 *2*，其产物是 *E. coli* 聚合酶抑制剂。宿主

表 16－6　T7 噬菌体的转录的时序调节

转录阶段	时间	聚合酶	转录组别	主要基因及功能
早期转录	前 6 分钟	*E. coli* RNA pol	转录物Ⅰ	10 个，*0.3*：抑制宿主限制酶
				0.7 基因：磷酸化 *E. coli* RNA pol
				1：T7 RNA pol，单链肽
晚期转录	6－16 分钟	T7 RNA pol	转录物Ⅱ	14 个，2：抑制宿主 RNA pol；合成复制酶系结构蛋白的基因
	12 分钟后	T7 RNA pol	转录物Ⅲ	15 个：编码与噬菌体粒子结构与装配有关的基因

		−10 +1(RNA起始点)	
保守序列		TAATACGACTCACTATAGGGAGA	CA
Ⅱ组启动子			
φ1.1A	AACGCCAAAT	CAATACGACTCACTATAGAGGGA	CC
φ1.1B	TTCTTCCGGT	TAATACGACTCACTACAGGAGAA	CC
φ1.3	GGACTGGAAG	TAATACGACTCAGTATAGGGACA	AT
φ1.5	AGTTAACTGG	TAATACGACTCACTAAAGGAGGT	AC
φ1.6	TGGTCACGCT	TAATACGACTCACTAAAGGAGAC	AC
φ2.5	AGCACCGAAG	TAATACGACTCACTATTAGGGAA	GA
φ3.8	CGTGGATAAT	TAATTGAACTCACTAAAGGGAGA	CC
φ4C	CCGACTGAGA	CAATCCGACTCACTAAAGAGAGA	GA
φ4.3	CGTCCCATTC	TAATACGACTCACTAAAGGAGAC	AC
φ4.7	TTCATGAATA	CTATTCGACTCACTATAGGAGAT	AT
Ⅲ组启动子			
φ6.5	GTCCCTAAAT	TAATACGACTCACTATAGGGAGA	TA
φ9	GCCGGGAATT	TAATACGACTCACTATAGGGAGA	CC
φ10	ACTTCGAAAT	TAATACGACTCACTATAGGGAGA	CC
φ17	GCGTAGGAAA	TAATACGACTCACTATAGGGAGA	GG
复制启动子			
φOL	TTGTCTTTAT	TAATACAACTCACTATAAGGAGA	GA
φOR	CACGATAAAT	TAATACGACTCACTATAGGGAGA	GG

图 16-38 T7 RNA pol 所识别的启动子核甘酸序列。(引自 J.D.Watson *et al.*: Molecular Biology of the Gene, 4th ed., 1987, Fig. 17-11)

的 RNA 聚合酶先受到基因 *0.7* 产物的磷酸化，再经抑制完全失去了作用，此时 T7 已不再需要 I 组的转录物了，因此废除了宿主 RNA 聚合酶的功能，而将自己编码的 RNA 聚合酶取而代之。第Ⅱ组的基因多与 T7 的复制酶系的合成及几种结构蛋白有关，但第Ⅱ组的启动子和复制的 φL 启动子因其保守序列中分别有 2～7 个碱基发生了改变，所以是较弱的启动子，但由于其位置排在 Ⅲ 组启动子的前面，所以先进入宿主得以转录，而不与第 Ⅲ 组启动子的竞争。

晚期转录的另一组基因就是第 Ⅲ 组转录物，它排在最后，故在感染 12 分钟后才得以表达。此转录物约含 15 个基因，它们主要编码头部蛋白，尾部蛋白以及负责装配，这样使得 T7 噬菌体不至于过早地装配成粒子。这一组虽排在最后，但启动子的结构和保守序列完全一致所以是极强的启动子，从而保证最后阶段结构蛋白的合成。

第Ⅰ组转录的终止位置在邻近 1.4 基因的上游区（1.3 和 1.4 之间），此是 *E. coli* RNA 聚合酶的终止子，它解离后 T7RNA 聚合酶则重新从第Ⅱ组起始转录。第Ⅱ组和第 Ⅲ 组都使用终止子 Tφ，由于自注入宿主缓慢，在第Ⅱ组转录时第 Ⅲ 组基因尚未注入，仍得不到表达。

Tφ 的终止效率是 90%，尚有 10% 可以延长转录 T7 的末端，Tφ 位于基因 *10* 和 11 之间。基因 10 是主要的头部蛋白，需要量大，它排在第Ⅱ组末是合理的，即有强启动子的启动，又在转录后被终止，这样不仅满足了装配颗粒的需要，又不至造成浪费。而排在基因 *10* 后面的基因都是编码少量头部蛋白和尾部蛋白的，无需大量的转录，后面越过 Tφ 延长转录的强度就可满足需要了，这本身也是一种终止子对转录的调控。

第五节　DNA 重排对转录起始的调控

真核 DNA 重排调节转录，产生不同的产物的例子有免疫球蛋白以及酵母交配型转换，而在原核生物中也有通过 DNA 交替的重排来调节转录的情况，最有名的例子就 Mu 噬菌体的重排以及沙门氏细菌的相转变。

在沙门氏菌（*S. typhimrium*）的鞭毛合成中，通过启动子方向的改变来调节不同鞭毛蛋白的合成。

细菌是通过摆动其鞭毛来运动的，许多沙门氏菌因具有 2 个非等位基控制鞭毛蛋白（鞭毛蛋白的亚基）的产生而出现两相性（diphasic）。同一菌落既可以表达为 H1 型[细菌处于 1 相（phase1）]，也可以表达为 H2 型[细菌处于 2 相（phase2）]。在细菌的分裂中有 1/1000 的概率会出现由一相转变为另一相的情况，就称为**相转变**（phase variation）。

负责合成两种鞭毛的基因位于不同的染色体座位。图 16－39 表明鞭毛蛋白的合成的循环调控。*H2* 基因和另一个 *rh1* 基因紧密连锁，*rh1* 基因编码 *H1* 阻遏物。这两个基因协调表达。处于 2 相时，在 *H2* 基因表达的同时，阻遏物基因也得到表达，阻止了 *H1* 基因的合成；在 1 相时，*H2* 基因和 *rh1* 基因都不表达，这样 *H1* 基因就进行合成。在此控制途径中，细菌的相取决于 *H2－rh1* 转录单位点是否有活性。

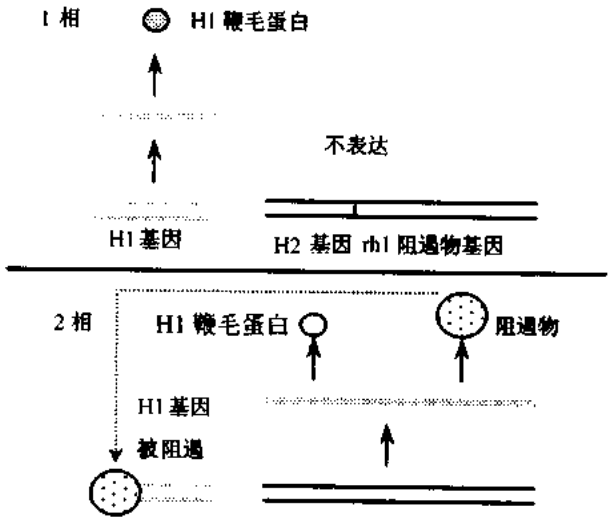


图 16－39　通过 H2 转录单位的活性来决定沙门氏菌。（仿 B. Lewin：Genes Ⅵ，1997，Fig. 32－25）

这个转录单位的活性是由与它相邻接的一个 DNA 片段来控制的，此片段长 995bp，两端是长 14bp 的反向重复序列（TRL 和 TRR）。*H2* 的起始密码子在反向重复序列 TRR 右侧 16bp 处。

hin 含有基因的 DNA 片段在 TRL（左反向重复序列）和 TRR（右反向重复序列之间，其产物 Hin（H segment inversion）蛋白通过反向重复序列之间的交互重组来介导整个片段的倒位（图 16－40）。*hin* 基因突变会使倒位的频率降低 10^4 倍。

H2－rh1 转录单位的启动子位于倒位片段之中，启动和转录单位方向相同时，转录在启动子处起始，而且持续通过 *H2－rh1*，导致 2 相的表达。当 *hin* 片段倒位时启动子和转录单位方向不同，转录单位不能表达，从而导致了 1 相表达。

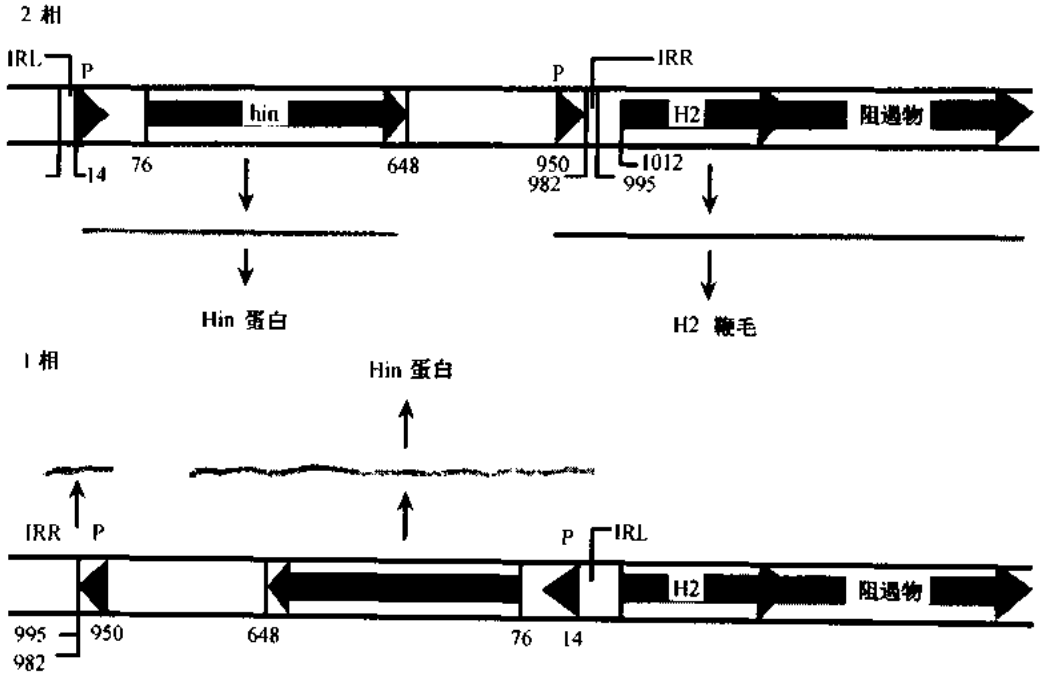


图 16－40　通过 995bp 的倒位区来控制 H2 鞭毛基因和阻遏物的表达。(仿 B. Lewin :Genes Ⅳ, 1992, Fig. 32－25)

在 Mu 噬菌体中的相关系统是在其基因组右端，显示出一种异乎寻常的特点。在右端有 3kb 的区域叫做 "G" 或倒位片段 (invertable segment)，在不同的 Mu 噬菌体的 DNA 中，这段序列方向不同。在感染 *E. coli* K^{12} 品系的 Mu 噬菌体中，G 片段的方向称为 G(＋)。经过诱导 Mu 噬菌体可能含有方向相反的 G 片段，此称为 G(－)。有一个 Mu 基因叫做 *gin* (G segment inversion)，正好位于 G 片段的右侧，是倒位反应所必须的 (图 16－41)。

G 片段含有的基因编码与噬菌体附着有关的蛋白。G 片段的方向控制着这些基因的表达。噬菌体 G 片段方向决定了宿主特异性。

图 16－42 表明 Mu 噬菌体 G 片段倒位能产生 4 种不同的尾丝蛋白。 G(＋) 方向

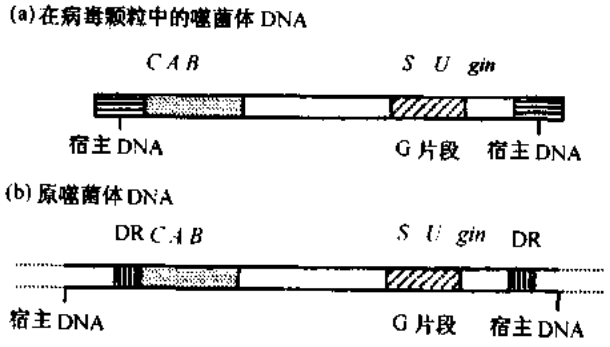


图 16－41　Mu 原噬菌体和游离噬菌体的结构。

时，S 和 U 基因表达，产物的分子量分别为 56kDa 和 21kDa，这些蛋白可使噬菌体吸附 *E.coli* K12 菌株，而不吸附 *E.coli* C。G(－) 方向时与此相反，基因 S′和 U′表达，它们产物的分子量分别为 48kDa 和 26kDa，这些蛋白允许噬菌体吸附 *E.coli* C，而不吸附*E.coli* K12。

这两类基因编码序列在两条互补链上，但方向相反。它们由 G 片段左侧的同一个

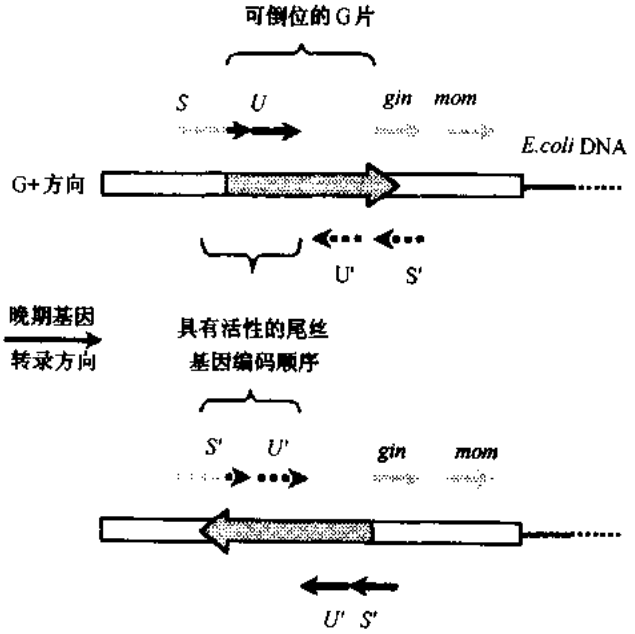


图 16－42　Mu 噬菌体的 G 片断倒位能产生 4 种不同的尾丝蛋白。
(参考 J.D.Watson et al:Molecular Biology of the Gene,4th ed.,1987,Fig.17－27)

启动子转录。两种方向的编码区都需 4kb，而 G 片段本身仅 3kb，因此还有 1kb 与启动子相联，在 G 片段上游邻接的 Sc 区(S 的恒定区)，它是 S 和 S′蛋白 N 端共有的序列。倒位的部分提供了 S 蛋白的可变区(Sv 和 Sv′)。

沙门氏菌中的 *hin* 基因和 Mu 噬菌体的 *gin* 基因能以互补的方式相互取代。在 *E.coli* 中也发现了相关的基因叫做 *pin*，它能和 Mu 噬菌体中 *gin* 的突变发生互补的作用。它们催化附近的 1.8 kb 的片段倒位。此片段的两端有 29bp 的反向重复序列。我们还不清楚此反应有什么功能。在 P1 噬菌体中还有一个 *cin* 基因负责 C 片段的倒位。

gin 和 *hin* 介导的倒位反应可在体外进行，但还需要一种尚未鉴别的宿主蛋白，反应的底物必须是超螺旋。*hin* 和 *gin* 反应需要另一种序列作为重组位点。此附加序列长 60bp，除了不能与重组位点本身靠得太近以外，可以位于底物分子的任何位置。

第六节　翻译的调控

翻译的几个主要的步骤：起始，延伸和终止都有不同方式的调控，以达到某种产物的平衡，既满足细胞的需要，同时又不造成浪费。

一、翻译起始的调控

1. 重叠核苷酸与翻译调控

在原核生物常常出现操纵子中相邻的基因有少量的DNA序列发生重叠，这不仅可以充分利用有限的碱基，而且有时可以起到调控的作用。下面举两个例子来说明这个问题。

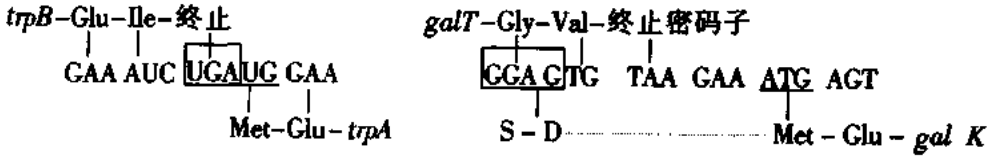
(1)色氨酸操纵子

色氨酸操纵子在前面已作了介绍，操纵子的5个基因中(图16-15) *E* 和 *D* 分别编码邻氨基苯甲酸合成酶的不同亚基组成四聚体。毫无疑问 *E*、*D* 产物的量一定要保持一种严格的当量关系，否则就造成浪费；同样 *A* 和 *B* 基因也是分别编码色氨酸合成酶的 α 亚基和 β 亚基，产量也要求一致。如何来保证这产量的一致性呢？仅仅依靠操纵子的调节还是不够的，因为往往上游的基因表达的量会大于下游基因。

人们发现当 *E* 发生突变时 *D* 的产量也随下降，但不影响下游的 *C* 基因，这是否是极性效应所致呢？回答是否定的。因为在这种情况下即使是 ρ 因子发生突变，*D* 的产量仍不能恢复。在上游基因发生无义突变时，有义密码子突变成了终止密码子，细胞中无对应的 α-tRNA，核糖体在此解离，紧跟其后的 ρ 因子可以延着 mRNA 追上 RNA 聚合酶，从 DNA 上解离下停止了转录。因此上游的基因的突变就会影响到下游基因的产量。若 ρ 因子的基因也发生突变就不会使 RNA 转录半途中止。但 ρ 因子发生突变，*D* 产物不能恢复，说明 *D* 产量的下降不是极性突变而致。那么影响到 *D* 产量的机制是什么呢？原来在 *E*、*D* 两个基因之间存在着翻译偶联效应。*trpE* 的终止密码子和 *trpD* 的起始密码子相互重叠。

trpE - Thr - Phe - 终止密码子
ACU UUC UGA UG GCU
Met - Ala - *trpD*

当 *trpE* 翻译终止时，核糖体尚未来得及解离，已处于 *trpD* 的起始密码子上，使翻译偶联起来，这两个基因的表达彼此协调，而与下游邻近的 *C* 基因就不存在这种翻译偶联的关系。*C* 基因单独编码吲哚甘油磷酸合成酶，也无需和 *trpE*、*D* 保持严格的一致性。*trpB* 和 *trpA* 也同样以这种偶联的关系来协调一致。



在 *E.coli* 的 *gal* 操纵子中 *E.T.K* 三个基因，*T* 与 *K* 以另一种形式重叠来达到协调一致：*gal-T* 基因的终止密码不和 Gal 的起始密码子重叠，但位于 *galK* 基因的 D-S 序列与起始密码子之间，当 *galT* 翻译终止时，核糖仍牢固地结合在 mRNA 上，继续翻译，使 *galT* 和 *galK* 的翻译保持一致。而在上游的 *galE* 基因却不和 *galT* 偶联。这是

由于 UDP－Gal 既可作为碳源又是细胞壁的重要的构筑材料，在无外源 Gal 时，可通过 *galE* 的产物异构酶的催化将 DDP－Glu 转变为 UDP－Gal，因此 *galE* 需要独立行动，不需和操纵中另外两个基因偶联。

在噬菌体中也存在着上叙 DNA 重叠调控的现象，例如在 *1.7* 和 *1.8* 基因之间存在着－**UGAUG**－的重叠序列，负责终止 *1.7* 的翻译，同时又是 *1.8* 的起始密码；在 *0.5* 基因和 *0.6* 基因之间也存在着类似的序列－**UAAUG**－，终止前一个基因的翻译，起始后一基因的翻译，以此协调两个相关基因的产物。

2. 翻译水平的自我调控

翻译起始的调节存在一种和转录调控类似的形式；无论编码区的起始位点是否对核糖体有利，调节分子直接或间接地决定翻译起始。这种调节分子可能是蛋白质，也可能是 RNA。

翻译的控制有些发生在 mRNA 中，编码核糖体蛋白的操纵子具有自调节的能力。此操纵子为一组结构基因的调节作了安排。可以使操纵子中单个基因的表达产生差异。

有一种蛋白可以结合到 mRNA 的靶区域上阻止核糖体识别以达到阻遏的功能。表面看它们与阻遏蛋白结合到 DNA 阻止 RNA 聚合酶利用启动子有相似之处。图 16－43 表明这种机制的共同类型。在此调控机制中，调节蛋白还可以直接结合到含有 AUG 的起始密码子的序列上，阻断核糖体的结合。

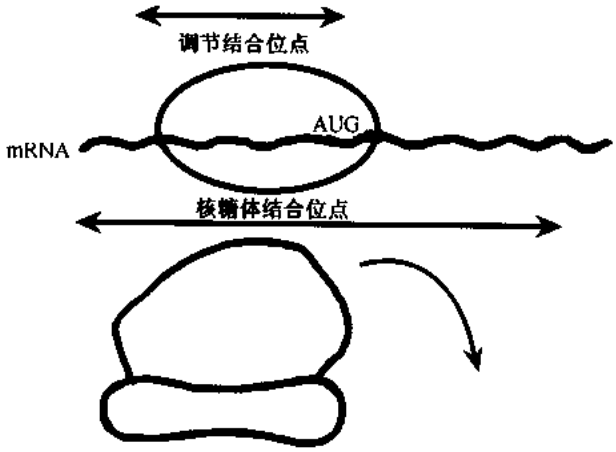


图 16－43　调节蛋白可通过结合 mRNA 上的核糖体结合位点上来阻断翻译。
（仿 B. Lewin：Genes Ⅵ，1997，Fig. 12－20）

在表 16－7 中翻译阻遏的例子及它们的靶基因，一类的例子是 RNA 噬菌体 R17 的外壳蛋白，它结合于含有核糖体结合位点的发夹回上；同样 T4 的 Reg A 蛋白结合在含有 AUG 起始密码子各种 T4 早期转录的 mRNA 的保守序列上；T4 DNA 聚合酶结合于特有的 mRNA 序列上，此序列含有核糖体结合所必需的 S－D 元件。约 70 种左右的蛋白构成细菌基因的表达系统，其中核糖体蛋白是主要成分，它与一些附属蛋白都和此系统有关。这些编码核糖体蛋白的基因，编码蛋白合成因子及 RNA 聚合酶亚基的基因，相互混合形成了一系列操纵子，这些蛋白基因在 *E. coli* 中只存在单个的拷贝。

协调控制保证了这些蛋白合成的量对于生长来说正好恰如其分。当细菌生长地较

快时，它们努力生产较大比例的基因表达装置。一种途径就是利用控制编码这些装置的基因的表达，且保证这些蛋白质的合成和 rRNAs 的水平保持适当的比例。

表 16－7　结合 mRNA 起始区的蛋白可以阻遏翻译

阻遏物	靶基因	作用位点包含
R17 外壳蛋白	R17 复制酶	包含核糖体结合位点的发夹回
T4 RegA	T4 早期 mRNA	含有起始密码子的各种顺序
T4 DNA 聚合酶	T4 DNA 聚合酶	S-D 顺序
T4 p32	基因 *32*	单链 5′前导顺序

B. Lewin: Genes Ⅵ, 1977, Table 12.2

图 16－44 表明了 6 个操纵子的组成。在 4 个紧密相连的操纵子中有一半基因是编码核糖体蛋白的(r-proteins)。这 4 个操纵子是 *str*、*spc*、*s10* 和 α。每个操纵子都以其第一个已鉴定的功能的基因来命名。*rif* 和 *L11* 操纵子在另一位点处。

每个操纵子都含有混合的功能。*str* 操纵子有核糖体小亚基的基因，还有 EF－G 和 EF-Tu 的基因；*spc* 和 *S10* 操纵子有核糖的大亚基和小基的基因，相间排列于其中；α 操纵子既有核糖体小亚基基因，又有 RNA 聚合酶 α 亚基的基因；*rif* 操纵子包含核糖体大亚基基因和 RNA 聚合酶的 β 与 β′亚基的基因。

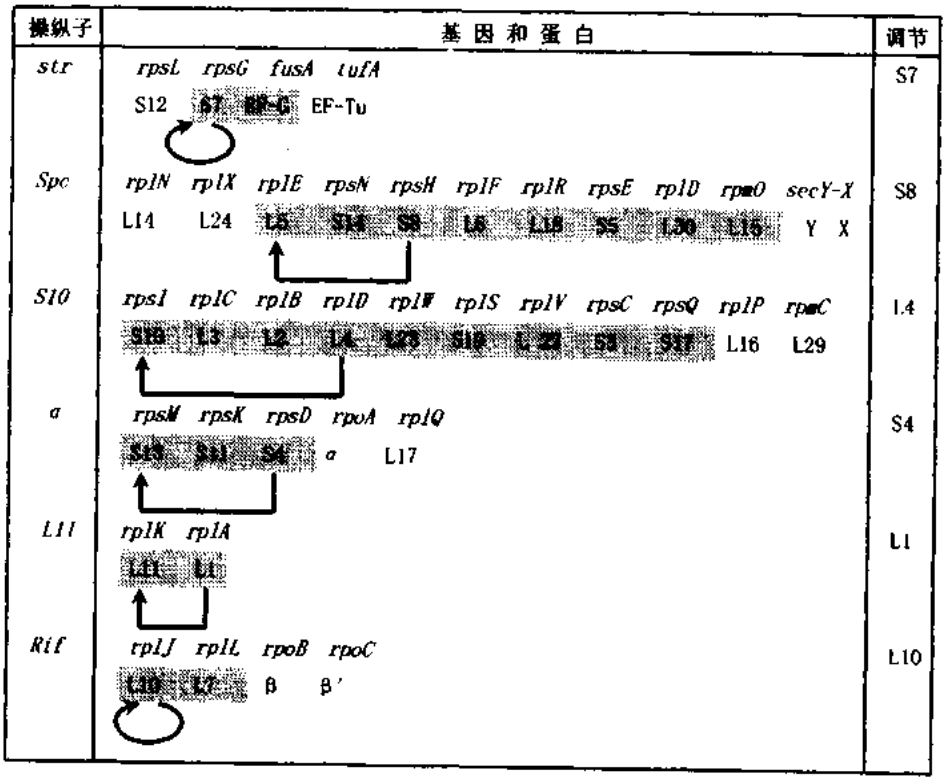


图 16－44　核糖体蛋白，蛋白合成因子及 RNA 聚合酶亚基的基因散布在几个自我调控的操纵子中，这些蛋白中大部分(灰色框)受调节蛋白的调控。(仿 B. Lewin: Genes Ⅵ, 1997, Fig12－32)

除一种核糖体蛋白以外其他的都需要保持与 rRNA 相适应的量。分散排列在不同

操纵子上的基因其产物的量需保持一致（如核糖体蛋白之间），而混合编排在一起的基因共产物又需要不同的量（如核糖体蛋白与聚合酶亚基以及延伸因子之间）。这就需一种有趣的协同调节。

图 16－45 中的 6 个操纵子的共同特点是某些基因受到一种产物的调节。在各种情况下产生调节蛋白的基因也是被调节的靶基因之一，这就是自我调控的例子。在 r－蛋白操纵子中，调节蛋白抑制了操纵子中邻近位点基因的表达，这是负自主调节的例子。

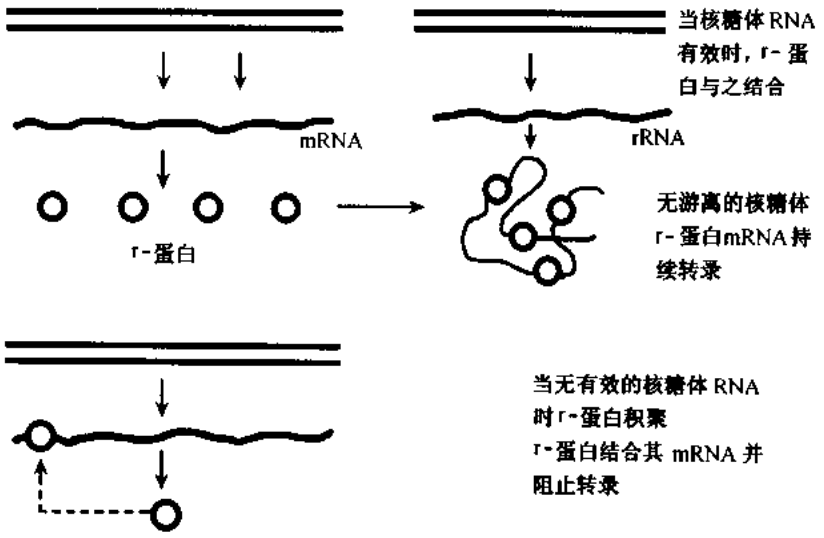


图 16－45　r－蛋白操纵子转录的自我控制并对 rRNA 的水平作出反应。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 17－33）

在多种情况下，蛋白质的积累抑制其本身的进一步合成，而且还涉及其他基因产物。这种作用常存在于多顺反子的翻译水平，而且常常能在体外重复这种现象。游离核糖体蛋白的存在可以触发翻译的阻遏。

每种调节蛋白都是可以直接与 rRNA 结合。它对翻译的作用是它能结合于作为自己模板的 mRNA 上。在 mRNA 上调节蛋白的这些结合位点也正是翻译的起始位点，或通过诱导构象的改变影响到有关的翻译起始的位点。如在 *S10* 和其后面一些基因的翻译。另外也存在调节蛋白的相互作用。就在上述操纵子的例子中，*L4* 也可以导致 *S10* 基因上游的非转录前导序列提前终止转录。

结合于 rRNA 的 r－蛋白所建立的自我调控表明存在一种 r－蛋白合成与 rRNA 合成连接的机制。图 16－46 表示 r－蛋白操纵子转录的自我调控模型。假设 r－蛋白在 rRNA 上的结合位点比在 mRNA 上的强。只要任何游离的 rRNA 是有效的，新合成的 r－蛋白将与之结合装配成核糖体。若无游离的 r－蛋白结合于 mRNA 上，翻译将持续。当一旦 rRNA 合成降低或停止，游离的 r－蛋白就开始积聚。然后它们有效地结合到它们自己的 mRMA 上，阻遏进一步的翻译。这种循环保证了每个 r－蛋白操纵子以相同的方式对 rRNA 的水平作出反应：一旦 r－蛋白相对于过剩，蛋白的合成就会停止。

这种调控模型完成了两个目标，首先是 r－蛋白的水平要和生长条件相适合。通过 rRNA 水平的控制，细胞控制了各种核糖体成分的生产。第二，由这些操纵子编码的另一些蛋白质是以它们特有的速率进行合成，与 r－蛋白基因的翻译相脱离。这样在

rif 操纵子中。RNA 聚合酶的 β 亚基受到了它们自身的自我调节，不由 L10 r－蛋白来控制。这就可以解释 RNA 聚合酶的量要比核糖体的成分少得多，它们在同一个操纵子中如何合成的问题。在这些操纵子的控制循环中，具有某种机制使一些协同调节的蛋白合成速率不同。有一种例外的核糖体蛋白是 L_7/L_{42} 在每个核糖体中有 4 个拷贝，它们（图 16－45）在转录时增加了效率。另一个例外是 EF-Tu 它存在的量大致和氨酰－tRNA 相等，约比核糖体多 10 倍左右，在这种情况下，有一个以上的基因以满足额外合成的需要。*tuf* 基因就分为 *tufA* 和 *tufB* 两个基因，它们是由彼此分离的循环来调节的，从而增加 EF-Tu 的合成。

T4 噬菌体的 32 基因在定量的基础上进行自我调节。p32 蛋白在遗传重组，DNA 配对和复制中起到举足轻重的作用，它的功能是凭借着它结合单链 DNA 的能力来实现的。无意突变可以导致蛋白失活使产量过剩。这样，当蛋白质失去功能时它产生得越多。这种效应发生在翻译水平。基因 32 的 mRNA 是稳定的，因而产量过剩是不考虑蛋白产物的行为。

图 16－46 表示基因 *32* 的调控模型，当在噬菌体感染的细胞中存在单链 DNA 的话，它容纳了 p32。但在缺乏单链 DNA 或者至少有过剩 p32 存在时，此蛋白就阻止其本身的 mRNA 的翻译。据推测，此作用是由 p32 直接地结合在 mRNA 上阻断了翻译的起始，有可能就结合在核糖体结合位点周围的 A－T 丰富区。

p32 在 mRNA 上的结合位点有两个特点，为循环控制效应提供了必需的条件。①p32 对于基因 *32* mRNA 上结合位点的亲和性必须要低于对单链 DNA 的亲和性，即结合 RNA 的平衡常数低于结合单链 DNA 的平衡常数。②p32 对其 mRNA 的亲和性必须大于其他 RNA。与 p32 结合的 mRNA 的一个重要特点是其调节区有一段序列具有二级结构。

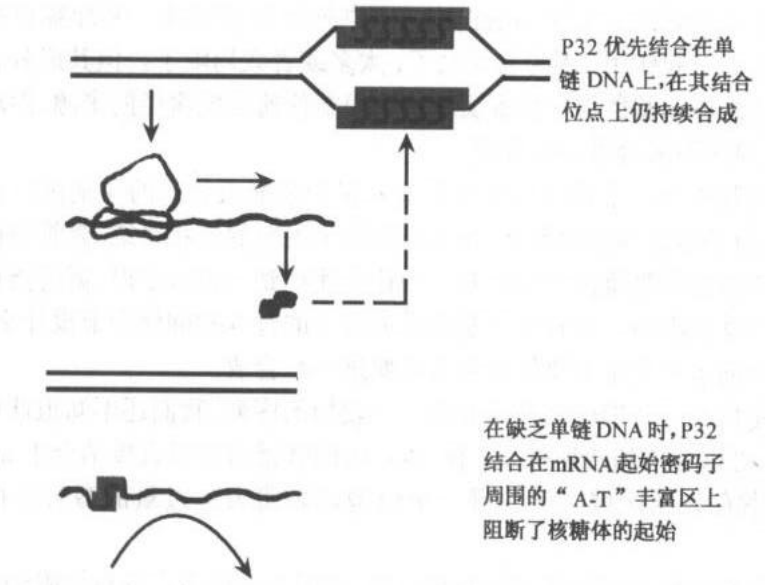


图 16－46 过量的 P^{32} 蛋白结合到它本身的 mRNA 上阻止核糖体起始翻译。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－34）

用已知的平衡常数我们可以在蛋白质浓度的情况下了解 p32 结合靶的位点。图 16－47 表明在蛋白浓度 $< 10^{-6}$mol/L 时，p32 结合于单链 DNA 上；浓度 $> 10^{-6}$mol/L 时，它结合于基因 *32* 的 mRNA 上；在更大浓度的情况下它可以和其他的 RNA 的序列结合。

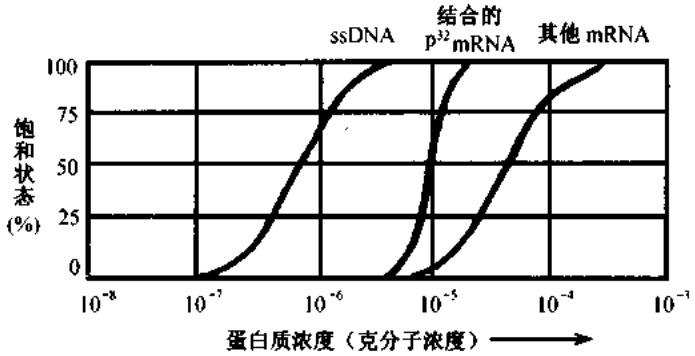


图 16－47　基因 32 蛋白对于不同底物的亲和性，它首先结合 ssDNA、然后结合其特有的 mRNA，再结合基他 mRNA。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－35）

这个结果表明 p32 的水平被控制在 $< 10^{-6}$mol/L，与约 2000 分子/细菌相协调。此与测定出的水平（1000～200 分子/细胞）相适应。

自动控制的特点是各种调节蛋白的相互作用是独特的：一种蛋白只作用于它本身的 mRNA 上，噬菌体 T4 提供了一个普遍性的翻译调节的例子，由 *regA* 基因编码的调节蛋白阻遏感染早期转录的各种基因的表达。RegA 蛋白是通过与 30S 亚基在 mRNA 起始位点上的竞争来阻断这些基因的 mRNA 的翻译。其作用直接地与一种阻遏蛋白的功能互相补充，而这种阻遏蛋白是可以与多种操纵子相结合的。

自主调节是在装配大分子的蛋白中一种共同的调节形式。装配颗粒本身并不适合作为一调节蛋白起作用，因为它太大了，太多或者太局限了。但其成分合成的需求可能在游离的前体亚基群体中得到反应。若阻断任何装配途径的来源，游离亚基就会积聚，并会关掉这种多余进一步合成。

在真核细胞中有一个共同的调节系统就属于装配大分子的亚基进行自我调节的类型，微管蛋白是形成微管的单体，是真核细胞中主要的丝状系统。微管蛋白 mRNA 的产生受到了游离的微管蛋白库的控制。当这个库达到一定浓度时，微管蛋白 mRNA 的进一步合成就受到阻碍。微管蛋白装配成大分子的过程和翻译调节没什么关系，是游离前体蛋白库的水平决定了单体是否有必要进一步合成。

调节的靶位点是在编码区开始处的一个较短的序列，我们还不知道此序列起什么作用，但我们可以考虑到两种模型。图 16－48 微管蛋白可以直接结合于 mRNA 上；或结合于新生肽存在的区域。无论哪一个模型都表明表达过量的微管蛋白可导致其 mRNA 降解。

自我控制与前面介绍的外部控制相对照，它属于一种内在的自我限制系统。一种调节蛋白结合在操纵子上，可以通过一种外来小分子的水平得到调节，这种小分子激活或抑制调节蛋白的活性。但在自我调节中关键性的参数是蛋白质本身的浓度。在细胞中游离蛋白库达到一定的分子浓度时，使 mRNA 不能成为蛋白合成的模板，从而终

止新蛋白的合成。

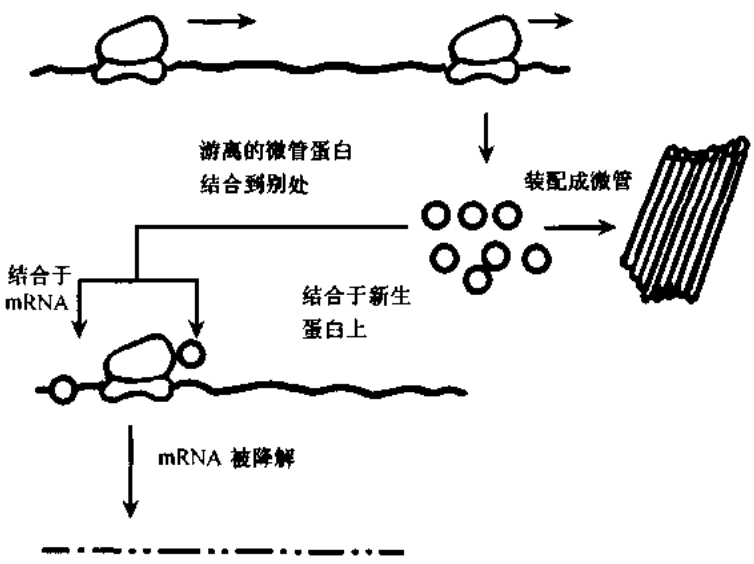


图 16-48 微管蛋白的自我调控。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12-36)

3. 二级结构和翻译的调控

E. coli 的 RNA 噬菌体 MS2, R17, f2 和 Qβ 都非常小(直径约为 250Å),是已知最简单的病毒。基因组长 3600~4200nt,只含 4 个基因:*A*、*cp*、*rep* 和 *lys*。Qβ 和 MS2 等不同,没有 *lys* 基因,但有 A_1 和 A_2 两个基因。*cp*(coat protein)外壳蛋白基因所编码的蛋白由 129 氨基酸构成,分子量为 13.7kDa。*A*(atachment)蛋白基因编码附着蛋白(含 393 氨基酸,分子量为 44kDa);*rep*(replicase)基因编码复制酶(含 544 氨基酸,分子量为 61kkDa);*lys*(lysis)蛋白基因编码裂解蛋白,它和 *cp*, *rep* 基因重叠,该蛋白含 75 氨基酸(图 16-49)。

A_2 编码成熟蛋白,它兼有裂解蛋白的功能;A1 蛋白是主要的病毒粒子蛋白,它是核糖体通读外壳蛋白终止密码子 UGA 而产生的。

一个 RNA 噬菌体基因组进入宿主细胞后,核糖体仅附着到 cp 基因开始的核糖体结合位点上,而不结合到 *A* 蛋白基因或 *rep* 基因的开始处,因为它们的核糖体结合位点被病毒 RNA 的二级结构所保护。相比这下,CP 蛋白的起始密码子 AUG 处被核糖体

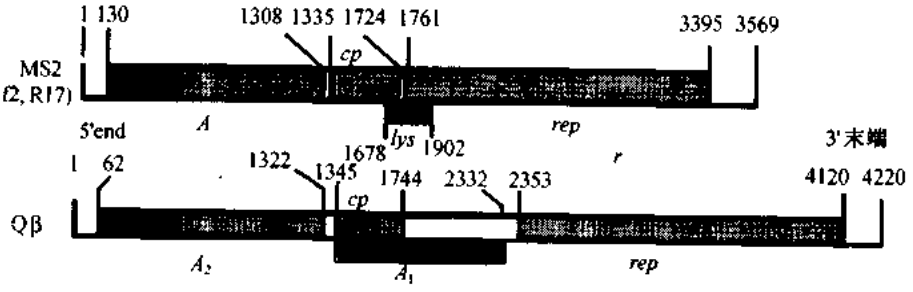
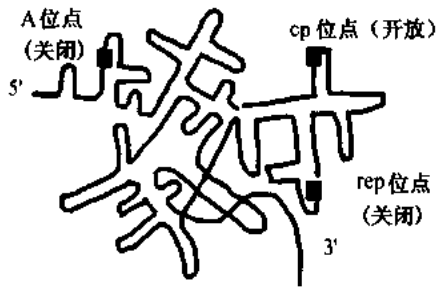
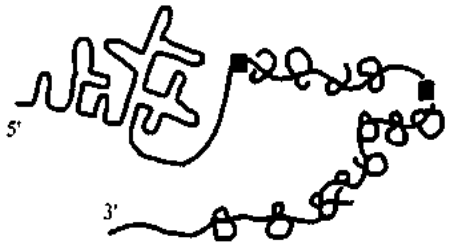


图 16-49 RNA 噬菌体 MS2(f2, R17)和 Qβ 的遗传图谱。A:附着蛋白;CP:外壳蛋白;Rep A_1:主要的病毒粒子蛋白;A_2:成熟蛋白。复制酶基因;Lys:溶菌酶基因。(引自 R. B. Winter and L. Gold: Cell. 1983; (33): 878)

所附着，因它暴露在二级结构的末端(图 16－50)。



(a) 在不合成蛋白时，“A”和“rep”位点是关闭的



(b) 核糖体越过 cp 基因，将封闭在二级结构中的 rep 位点打开，使其能有效地翻译

图 16－50　表明 RNA 噬菌体的 RNA 二级结构对翻译的调控。
(仿 J. D. Nafson: Molecular Biology of the Gene, 4th ed., Fig. 17－35)

由于 *A* 和 *rep* 两个位点被封闭在二级结构中，首先打开的是 *CP* 位点。当核糖体阅读到 *CP* 位点时，使形成二级结构的氢链断裂。随着翻译的进行，将其下游的 *rep* 位点也冲开了。这样 *rep* 基因总是依赖于位于前面的 *cp* 位点和核糖体的结合。通过发现 *cp* 基因第 6 个氨基酸上发生的无意突变是极性的，它阻止了 *rep* 基因的表达，尽管 *rep* 基因是完整的，这才首次提出这种依赖性。核糖停止在第 6 个氨基酸的密码子之后，就不能打开氨基酸的起始位点，而在 50, 54 或 70 位点上的无意突变并非是极性的，因此在核糖体是这么远时，*rep* 的起始序列就已打开了。

虽然 *rep* 的核糖体结合位点每次都被 *CP* 顺反子的翻译所打开，但是 *RNA* 噬菌体的 *CP* 蛋白多肽链产生的量要比复制酶多肽链多得多，这就意味着其间存在一种调控：新产生的外壳蛋白的亚基可以特异地附着到 *rep* 基因的核糖体结合位点，阻止了核糖体的附着。这样外壳蛋白就成了复制酶基因的特异翻译阻遏物。在感染 10 分钟后，外壳蛋白足以阻断复制酶的进一步合成(图 16－51)。

RNA 噬菌体的复制酶复合体是由本身合成的复制酶和宿主合成的 Tu，Ts 及 S_1 三种蛋白组成复合体。复制以原有的 RNA （＋链）为模板，由 5′到 3′合成新的负链（图 16－52），负链合成后再以它为模板合成正链。

A 蛋白的翻译是趁复制开始时，其位点裸露了出来，在尚未形成二级结构之前核糖体结合上去，进行翻译，产生 A 蛋白。

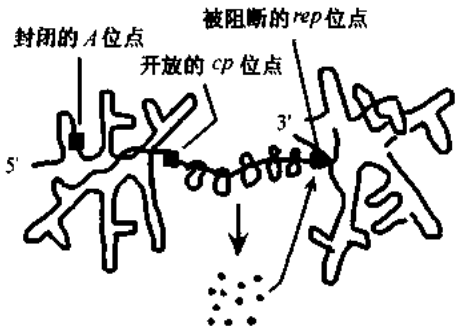


图 16－51　外壳蛋白分子结合到 *rep* 基因的起始位点上，阻遏其合成。(仿 J.D.Watson *et al*.:Molecular Biology of the Gene,4th ed.,1987, Fig.17－36)

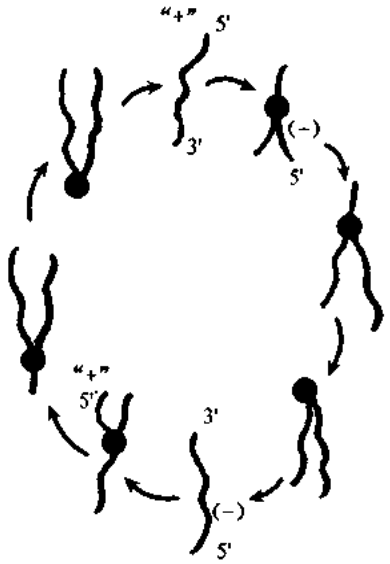


图 16－52　RNA 噬菌体的复制 (Qβ, MS2.R17.f2)。(仿 J.D.Watson *et al*.:Molecular Biology of the Gene,4th ed.,1987,Fig.17－39)

4. 反义 RNA 的调控

直到现在我们所讨论的反式作用调节因子都是蛋白质。用一个 RNA 作为调节物同样能形成调节网络。其实，最初的操纵子模型左边的调节物就考虑到可能是 RNA 或蛋白质。

现在有一些例子表明小分子 RNA 也可调节基因的表达。和蛋白质调节物一样，此 RNA 是独立合成的分子，与靶位点的特殊序列是分开的。调节物 RNA 的靶顺序是单链核苷酸顺序，调节物 RNA 的功能是和靶顺序互补，形成一个双链区。此调节物 RNA 的作用可能有两种机制。① 和靶核苷酸顺序形成双链区，直接阻碍其功能，如翻译的起始。②在靶分子的部分区域形成双链区，改变其他区域的构象，从而影响其功能。两种类型 RNA 介导的调节其共同特点是改变靶顺序的二级结构，控制其活性。

在 RNA 调节物和蛋白质调节物之间的差别是对于阻遏操纵子 RNA 不具有变构的性能，它不能通过改变识别靶的能力来对别的小分子作出反应。它能通过控制其基因的转录打开，或通过一种酶降解 RNA 调节产物而关闭。

在控制翻译的复制中意味着 RNA－RNA 的相互作用。小分子 RNA 直接控制 *E.coli ompF* 基因的翻译(图 16－53)。

1984 年 Mizuno 和 Iwoue 在研究 *E.coli* 外膜蛋白时发现了**干扰的 mRNA 的互补** RNA(mic-RNA, mRNA-interfering complementary RNA)。

在 *E.coli* 中有两种外膜蛋白 OmpC 和 OmpF，在渗透压增高时 OmpC 产量增加，在渗透压减小时 OmpC 产量受到控制；与此相反另一种膜蛋白 OmpF 在高渗透压条件下产量受控，而低渗时其产量增加。它们如何对渗透压的改变作出反应呢？*OmpC* 和 *Omp*F 两个基因编码了 OmpC 和 OmpF 两种外膜蛋白，这两个基因并不连锁，但却受到培养基的渗透压的控制。它们的表达和渗透压改变有关。*EnvZ* 基因编码一种作为渗透压感受器的一种受体蛋白。当渗透压增加时，EnvZ 激活 *OmpR* 的产物 (一种正调节蛋白)。它可以激活两个基因转录，那就是 *OmpF* 和调节蛋白 *mic F*，这两个基因相互连锁，但反向转录，调控区位于两个基因之间。

micF 的产物是一条长 174nt 的 RNA，称为 **micRNA**(mRNA－interfering－complementary mRNA)，即**干扰 mRNA 的互补 RNA**。一般都称之为**反义 RNA**(antisense RNA)。MicF RNA 可以和OmpF mRNA 上包含的核糖体结合位点翻译起始区互补结合，形成双链区。

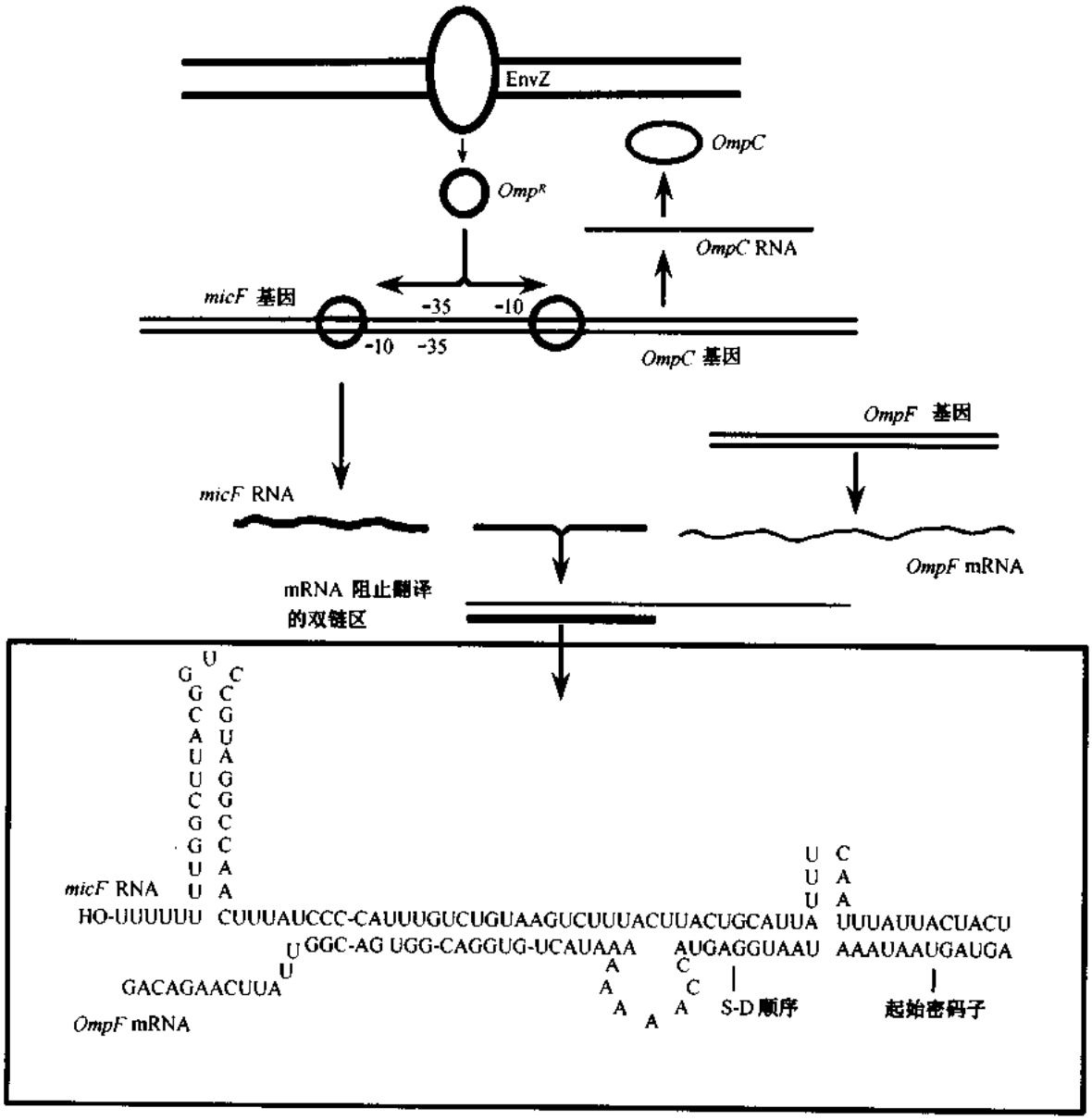


图 16－53　反义 RNA 的调控机理。(仿 B. Lewin: Genes Ⅴ, Ⅳ, 1994, 1992, Fig. 16－15)

MicF RNA通过和 OmpF mRNA 的结合来作为一种调节物并阻止其翻译。它也可能和 OmpF mRNA 形成不稳定的双链区，如用核酸酶作用双链区，结果是敏感的。当渗透压增加时会导致 micRNA 的合成，而关闭了 OmpF mRNA 的翻译。

看来似乎 micRNA 的合成足可以使其靶 RNA 失活。将编码 micRNA 的人工基因导入 *E. coli*，它们可以阻止特殊的靶基因的表达。在 λ 噬菌体的生活周期中也存在着这种反义调控(见第 17 章)。

反义基因可被导入到真核细胞中。方法是将靶基因反向插入重组载体的启动子后面，再导入细胞，这便称为**反义技术**（图 16－54）。当将胸苷激酶（tK）的反义基因导入细胞，它将抑制细胞本身的 *tK* 基因的合成。由于其作用和反义 RNA 的数量有关，因此常常并不能起到完全抑制的作用。但看来过量的反义 RNA 应能有效阻止靶顺序的翻译。实际上反义 RNA 不需和 mRNA 等长，只要靶 mRNA 的一小部分结合可达到

调节了。一般只要 < 100 碱基长的靶的 RNA5′区域的反义 RNA 就可以有效地抑制翻译。

反义技术抑制表达是在什么水平呢？原则上说它可能阻止一些基因的转录，RNA 的加工，或者 mRNA 的翻译。抑制作用往往依赖于 RNA－RNA 双链分子的形成。但抑制作用既能发生核中，也能发生在胞质中。在培养细胞中有义－反义 RNA 双链可以在核中形成，阻止正常的加工或者 RNA 的转录。另一种情况下是将反义 RNA 注入到细胞质中，通过和 mRNA 5′端区域形成双链来抑制翻译。

此技术常常是关闭某些基因的有效方法，例如导入反义 RNA 来研究基因的调控。这种技术的一种延伸就是把反义基因置于启动子的控制之下，研究其本身的调节。靶基因可能被关闭，也可能不受反义 RNA 产物的调节。反义技术在应用上有着广阔的前景。现已应用于抗肿瘤，抗病毒的治疗研究，对某些遗传病甚至寄生虫病的治

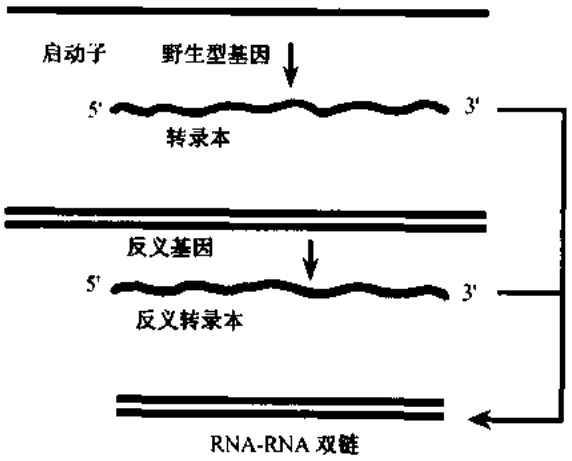


图 16－54　反义 RNA 和靶 RNA 互补结合阻止其翻译。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12－45）

疗（如锥虫病）也可能发挥作用。农业上除了培养抗病毒植株外，现已应用于果蔬运输的保鲜以及珍奇花卉的培育。

二、翻译的延伸调控

1. 稀有密码子的调控

在原核生物中，有时同一个操纵子中的基因其功能并无相关性，如前面介绍的核糖体蛋白的一些基因的操纵等，那么它们的产量就不可能要求一致，但又同在一个操纵子中，如何来进行调节呢？除自我调控外还可以利用别的途径，如利用稀有密码子进行调控就是有效的方法之一。

所谓的稀有密码子是在一般的编码中利用频率很低的密码子。如在 25 种非调控蛋白中 Ile 的几种密码子频率如表 16－8 所示，其中 AUA 使用频率仅有 1%，它就是一种稀有的密码子。这就意味着在一般情况下，细菌中相应于 AUA 的 tRNA 比较少。

现在再来看一下 *E. coli* 的 *rpsU* 操纵子，由 *rpsU*，*dnaG* 和 *rpoD* 三个基因组成簇（图 16－55）。*rps* 指令合成核糖体小亚基蛋白 *S21*，在每个细胞中有 4 万个分子，产量很高。与其相邻的 *dnaG* 基因编码引物酶，催化合成冈崎片段的引物，每个细胞中含50

表 16－8　在不同产物中密码子使用频率不同

密码子	在 25 种非调控蛋白中(%)	在引物酶中(%)	在 σ 因子中(%)
AUU	37	36	26
AUC	62	32	74
AUA	1	32	0

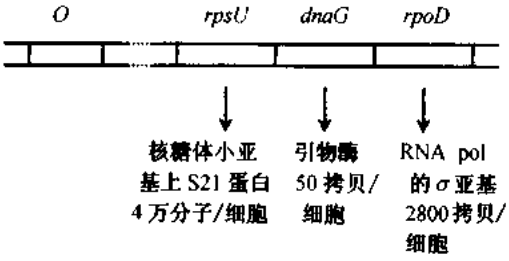


图 16－55　*rpsU* 操纵子中 3 个基因产物的拷贝数完全不同。

拷贝，相对的产量是很低的。*rpoD* 编码 RNA 聚合酶的 σ 亚基，每个细胞中有 2800 个拷贝，产量界于前二者之间。这三个基因连锁在同一操纵子上，为什么它们产物的量差异如此之大呢，表中表明编码 Ile 的 AUA 在一般情况使用频率仅有 1%，而在 *dnaG* 基因中它占有的频率是 32%几乎占了 1/ 3，但细胞中其相应的 tRNA 含量很少，不易获得，这样就延长了核糖体在 mRNA 上行经的时间，从而降低了翻译的速度。*rpoD* 基因正好与此相反 AUA 的分布频率为 0。根本就不使用这个稀有密码子，所以合成 σ 亚基不受其影响，而使用最多的密码子 AUC 其 tRNA 也是细胞中含量最丰富的，正好可以满足其需要，因此其产量可以高达 2800 个拷贝。

2. 二级结构对翻译的调控

衰减作用是原核生物利用 RNA 的二级结构对转录进行调控。利用相似的机制还可以对翻译进行调控，较典型的例子就是抗红霉素基因利用 mRNA 二级结构的改变来调控甲基化酶的表达，以对环境中是否存在红霉素作出反应(图 16－56)。抗红霉素基因的 mRNA 的前导顺序中有 4 段 RNA 顺序，就像衰减子一样，可以相互配对形成二级结构。当无红霉素存在时，核糖体顺利地翻译了前导肽，并从 mRNA 上释放下来，使得 1－2，3－4 相互配对，形成了二级结构。而其编码甲基化酶区域的 S－D 顺序正好处在 3－4 区之间，这样被二级结构所包围，无法在翻译起始中发挥作用。因此甲基化酶不能产生，此时的 mRNA 呈现的是关闭构象。当红霉素存在时，它能与核糖体大亚基的 23S rRNA 结合，抑制了前导肽的翻译，因此核糖体停留在 mRNA 的 1 区，使 2 区和 3 区得以相互配对形成发夹结构，这样甲基化酶的 S－D 顺序就游离出来可以和第二个核糖体结合，翻译出甲基化酶。此酶的作用是使 23S rRNA 上的腺苷酸甲基化，结果红霉素就无法与其结合，从而达到抗红霉素的作用。

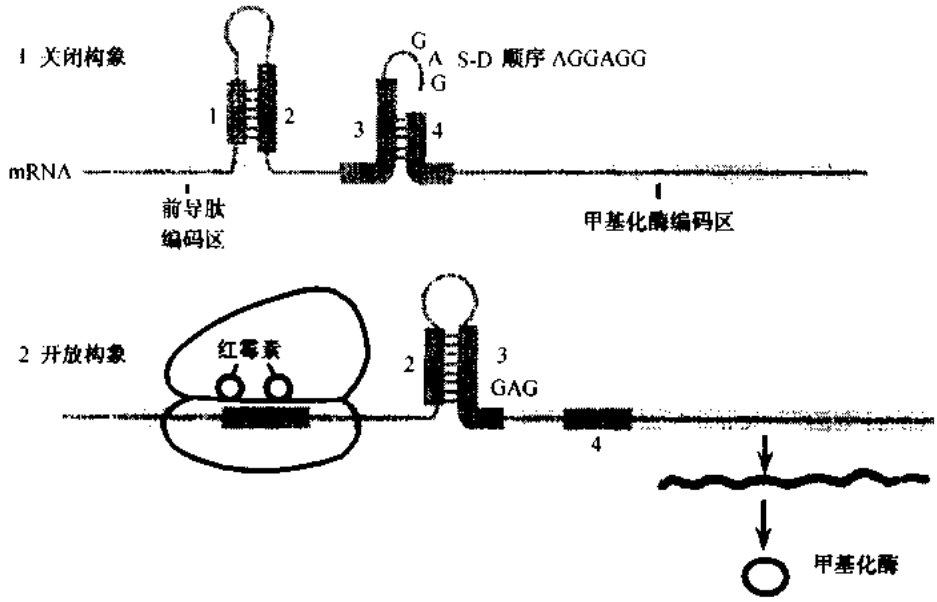


图 16-56 红霉素抗性基因 mRNA 二级结构的改变对甲基化酶的翻译调控。
(仿 J.D.Watson *et al.*: Molecular Biology of the Gene, 4th ed., 1987, Fig. 16-25)

三、翻译的终止调控

1. 严紧反应

当细菌发现它们自己生长在饥饿的条件下，缺乏维持蛋白质合成的氨基酸时，它们将大部分活性区域都关闭掉。这就称为**严紧反应**(strigent response)，这是它们抵御不良条件，保存自己的一种机制。细菌通过仅仅维持最低量的活性来节约其资源，直到条件改善时，它们又恢复活动，所有代谢区域也都活跃起来。

严紧反应导致 rRNA 和 tRNA 合成大量减少(10～20 倍)，使 RNA 的总量下降到正常水平的 5%～10%，部分种类的 mRNA 的减少，导致 mRNA 总合成量减少约 3 倍。而蛋白质降解的速度增加，很多代谢进行调整，显然核苷酸，碳水化合物，肽类等的合成都随之减少。

严紧反应导致两种特殊核苷酸积聚：①ppGpp—四磷酸鸟苷(在 G 的 5′和 3′位点各附着两个磷酸)；②pppGpp—五磷酸鸟苷(鸟苷 5′－三磷酸～3′二磷酸)。

人们最初发现细菌在氨基酸饥饿时，出现两种特殊的核苷酸，其电泳的迁移率和一般的核酸不同，感到很奇怪，就称之为“魔斑Ⅰ”和“魔斑Ⅱ”，后来发现魔斑Ⅰ便是 ppGpp，魔斑Ⅱ是 pppGpp。

这些鸟苷是典型的小分子效应物，预期它们的功能是和靶蛋白结合改变它们的活性，有时它们被称为(p)ppGpp。(p)ppGpp 的功能是调节细胞活性大分子的协调性。它们的产物是由两种途径来控制的。在严峻的条件下严紧反应可以触发 (p) ppGpp 的增加。一些未知因素的调节也会导致(p)ppGpp 水平与细菌生长速度之间的反向协调。

任何一种氨基酸的缺乏或使任何氨基酰－tRNA 合成酶的失活突变都足以起始严紧反应，此表明严紧反应的触发器是位于核糖体 A 位点中的空载 tRNA。当然在正常条

件下仅有氨基酸 - tRNA 在 EE - Tu 的作用下位于 A 位点。但当氨基酸 - tRNA 对一个特殊的密码子不能作出有效反应时，空载 tRNA 便能得以进入，当然这就阻断了核糖体的进程，而触发了一个**空转反应**(idiling reaction)。

通过空转反应产生 (p) ppGpp 的有关成分已通过**松弛型突变** (relaxed(*rel*) mutants) 被鉴别出来。*rel* 突变能去除严紧反应。这样氨基酸的饥饿便不会导致合成任何稳定的 RNA 或者通常见到的各种其他的反应。

松弛突变大部分基因位点位于 *relA* 基因中，此基因编码一种蛋白，称为**严紧因子** (stringent factor)。此因子和核糖体结合，虽然它的总量较低 - 每 200 核糖体低于 1 分子的严紧因子。因此似乎只有在核糖体少量存在时才能产生严紧反应。

从严紧反应中分离的核糖体在体外能合成 ppGpp 和 pppGpp。A 位点是通过一个空载 tRNA 对密码子的特异反应所提供的。从松弛突变体中提取的核糖体不能执行严紧反应，但若有严紧因子存在它们是能合成(p)ppGpp 的。

图 16 - 57 表示 (p) ppGpp 的合成途径，严紧因子 (RelA) 是一种 (p) ppGpp 的合成酶，它可以催化 ATP 将焦磷酸加到另一个 GTP 或 GDD 的 3′位点。

当条件恢复到正常时，ppGpp 怎样被去除的呢？有一个基因叫做 ***spoT***，它编码一种酶，主要的作用是催化 ppGpp 的降解。这种酶的活性导致 ppGpp 迅速降解其半衰期为 20 秒左右，因此 (p) ppGpp 合成结束时，严紧反应很快就消除。*spoT* 突变会提高 ppGpp 的水平，并会慢慢增加。

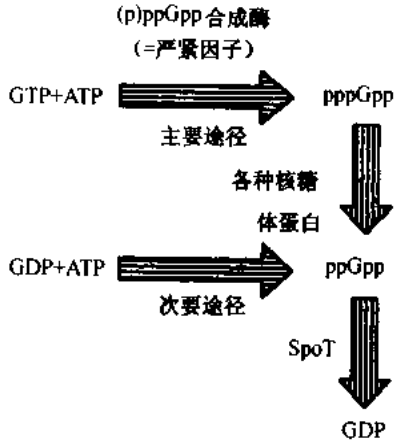


图 16 - 57　严紧因子催化(p)ppGpp 的合成。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 15 - 31)

RelA 酶用 GTP 作为底物的频率是较高的，所以 pppGpp 的产生是占优势的。但 pppGpp 可以通过各种酶转化为 ppGpp，其中翻译因子 EF - Tu 和 EF - G 是可以去磷酸化的。通过 pppGpp 产生 ppGpp 是最通用的路线，而 ppGpp 通常就是严紧反应的效应物。

图 16 - 58 表明核糖体对空载 tRNA 进入的反应与正常蛋白质合成的比较。当 ET - Tu 将氨基酸 tRNA 放在 A 位点时肽链随着核糖的移动而合成；但是当空载 tRNA 在 A 位点与密码子配对时，核糖体仍保持不动，并进行了空转反应。

提纯的 RelA 酶它本身是没有活性的。而在核糖体存在时它才有活性。它的活性是由核糖体在合成蛋白中的状态所控制的。此控制的特点通过另一个位点的松弛突变而显示，此突变原来称为 *relC*，现在弄清楚它就是编码 50S 亚基 L11 蛋白的 *rplK* 基因。

此蛋白位于 A 位点和 P 位点的附近，它在此位置对合适的配对作出反应，空载

tRNA 是在 A 位。L11 蛋白或某些其他成分构象的改变可能激活 RelA 酶，这样空转反应就代替了肽酰 – tRNA 的转位。

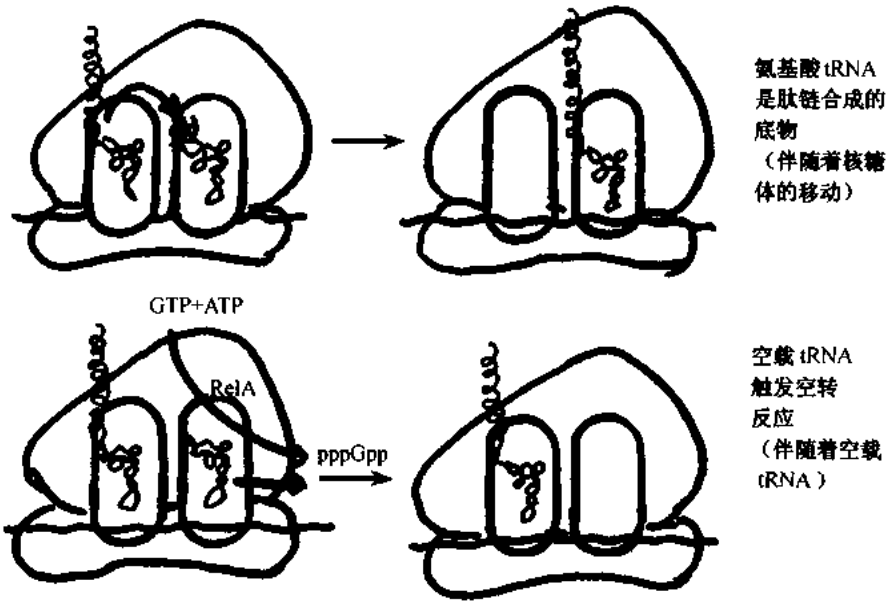


图 16 – 58　一正常蛋白质的合成和空转反应的比较。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 12 – 29）

(p) ppGpp 合成的每条途径都触发空载 tRNA 从 A 位点释放出来，因此 (p) ppGpp 的合成是一种对空载 tRNA 水平的持续反应。在饥饿条件下，当氨基酰 – tRNA 不能对 A 位点的密码子作出有效反应时，核糖体便停滞不前。空载 tRNA 的进入，触发了 (p) ppGpp 分子的合成。并将空载 tRNA 排出，使 A 位重新空出来。核糖体是恢复多肽的合成，还是进行另一轮的空转反应，关键是取决于氨基酰 – tRNA 是否有效。

ppGpp 有什么作用呢？它是一系列反应的效应物，包括抑制转录。许多反应已被报导，其中 2 个较为突出：①rRNA 操纵子的启动子其转录起始被严紧反应特异抑制了。严紧调节的启动子发生突变能消除严紧控制，表明此效应需要和特异启动子顺序相互作用；②很多或大部分模板的转录延伸阶段被 ppGpp 缩短了，此是 RNA 聚合酶停顿的增加而引起的。此效应反应了在细胞内加入 ppGpp 时转录效率普遍下降。现在尚不清楚这种抑制的特异性，若不同的操纵子之间这种抑制的变化很大，某些操纵子抑制作用更为显著的活，也是不足为奇的。

异常核苷酸在两种控制系统中都能发挥作用是很有趣的现象。这两种途径的抑制对细菌来说是特异的。在营养缺乏时鸟苷触发了严紧反应。而在缺乏葡萄糖时，cAMP 触发碳源利用的转换。

2. mRNA 稳定性对翻译的调控

在 *E. coli* 中尚未发现 5′ – 3′外切活性的核糖核酸酶，降解 mRNA 的酶有两种：RNaseII 和多核苷酸磷酸化酶，这两种酶都是 3′ – 5′的外切酶，但 mRNA 的二级结构可以阻遏这些酶的作用。

在 *E. coli* 和沙门氏菌中，发现一种高度保守的反向重复顺序 (IR)，对 mRNA 的稳

定性起着重要的作用。在 *E.coli* 中这种 IR 估计含有 500～1000 拷贝。它们有的位于 3′端非编码区，有的在基因间的间隔区。IR 的存在提供了形成茎环结构的可能性，从而增加 mRNA 上游部分的半衰期，对下游部分影响不大。这是由于 IR 的存在可以防止 3′－5′外切酶的降解作用。因此在多顺反子的操纵子中，基因间的 IR 可以特异地使其某些基因上游 mRNA 得到保护。例如在 *E.coli* 的麦芽糖操纵子中的 *malE* 和 *malF* 基因之间存在 2 个 IR 顺序（图 16－59）。*malE* 和 *malF* 虽然同在一个操纵子中，而且紧密连锁，但 *malE* 的产物（周质结合蛋白）要比 *malF* 的产物（一种 40kDa 的内膜蛋白）的含量高 20～40 倍。这可能由于 *malG* 和 *malF* 的 mRNA 区域不如 *malE* 的区域稳定，在 *malE* 3′端有 2 个 IR 存在，可以形成茎环保护其不被外切酶所降解。若 IR 区缺失，那么 *malE* 产物的量就会减少 9 倍。

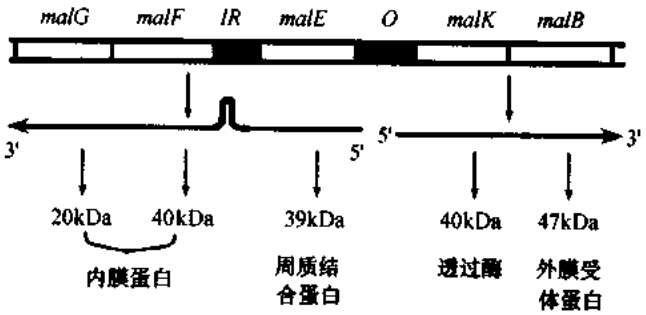


图 16－59　麦芽糖操纵子和 IR 的位置。

第十七章

噬菌体的裂解途径和溶原化途径的选择和调控

基本概念

λ噬菌体的调控有多种形式，有正调节、负调节、自主性的反馈调节、抗终止调节、反义调节及反向调节等。其机制之精巧，过程之复杂，形式之多样堪为集原核生物调控之大成。

C Ⅰ蛋白是λ噬菌体的阻遏蛋白，由两个亚基组成，每个亚基都有两个功能区，N端与操纵原体结合，C端具有相互结合成二聚体的能力。C Ⅰ蛋白的二聚体与二聚体之间有协同效应。

C Ⅰ蛋白具有阻遏 P_L，P_R 启动子的功能，若发生突变会导致宿主的裂解。C Ⅰ 不仅阻止已经溶原化的λ原噬菌体基因的表达，也使新进入溶原化细菌中的λ噬菌体不能复制，从而使宿主产生"免疫"性。

C Ⅰ和 Lac Ⅰ相似具有反式作同，阻遏 P_R 和 PL，是一种负调节蛋白。

λ噬菌体的操纵基因 O_R 和 O_L 各有3个能和阻遏物结合的位点 O_{R1}，O_{R2}，O_{R3}，O_{L1}，O_{L2}，O_{L3}。每个位点大约 17bp。

3个位点和阻遏物的亲和力各不相同，这对裂解和溶原途径的选择起着重要的作用。

17bp 序列构成了不同的回文结构，回文的对称性和亲和力的大小有关。

C Ⅰ的N端以 H－T－H 的形式与 DNA 的大沟结合。

Cro 蛋白也是一种阻遏蛋白，以二聚体的形式与操纵基因结合，但 Cro 最易和 O_{R3} 结合，而 C Ⅰ最容易与 O_{R1}，O_{R2}。结合 Cro 无协同效应。

C Ⅰ蛋白对 *Cro* 基因的转录起负控制的作用，而对 *c Ⅰ* 基因本身的转录起着正负控制双重作用。这是由于 *c Ⅰ* 的维持启动子 P_{RM} 和 P_R 部分重叠，当 C Ⅰ蛋白浓度高时与 O_{R1}，O_{R2}，O_{R3} 3个位点结合，同时也抑制了 P_{RM} 转录，起负调节作用；当 c Ⅰ蛋白浓度低时只和 O_{R1}，O_{R2} 结合，空出了 O_{R3}，促进 RNA 聚合酶与 P_{RM} 结合，起正调节的作用。

Cro 蛋白最易和 O_{R3} 结合，所以对 *cI* 基因和 *cro* 基因却只起负控制作用。

C Ⅱ蛋白能激活 C Ⅰ建立溶原化启动子 P_{RE}，整合酶启动子 P_I 以及反终止蛋白 Q 的反义 RNA 启动子 P_{AQ}，这些启动子的激活却对溶原化有利。

λ噬菌体也具有反义的 mRNA 调节的机制。一是 P_{RE} 启动子其转录方向与 *cro* 基因的转录方向相反，其 5′端和 *cro* 基因的 mRNA 重叠，可反义抑制 *cro* 的表达；二是 P_{AQ} 启动子转录一段长约 200b 的 RNA，不编码蛋白，而和抗终止蛋白 Q 的 mRNA 5′端互补，起到反义抑制的作用。

λ噬菌体有2个抗终止基因 *N* 和 *Q*，它们的产物可在一定的位置和 RNA 聚合酶结合，使其顺利通过早早期和晚早期转录的终止子，有利于宿主的裂解。

反向调节是λ噬菌体所特有的，利用不同启动子 P_L 和 P_I 所产生的 RNA，虽序列重叠，但长度不同，形成二级结构也不同，前者的 mRNA 易被 RNase Ⅲ剪切，继而被核酸酶所降解，从而不能表达，而只有 P_I 启动子转录的 *int* mRNA 才是稳定的可以翻译成整合酶，供溶原化时λ基因组插入到宿主染色体上所需。

λ噬菌体的遗传背景，生活周期都已研究得比较清楚，特别是调控，几乎各种调节途径在λ噬菌体中都存在。在基因工程中，λ噬菌体经改装后又常用作转染的载体，因此掌握λ噬菌体的遗传背景及调控机制是十分有用的，故单列一章系统地加以介绍。

第一节　λ噬菌体的基因组及早、晚期转录

一、裂解的级联调节

前面（见第 8 章）已经初步介绍了 λ 噬菌体的生活周期及转导。λ 噬菌体长 48502nt，共 61 个基因，其中 32 个较为重要（图 17－1，表 17－1）。λ 噬菌体提供了最为复杂的级联调控。在它发育中有两种可调控的连续阶段，溶原化周期和裂解周期，而这两个周期是相互联系的（图 17－2）。

当 λDNA 进入新的宿主细胞时裂解和溶原化的途径并存。**早早期**和**晚早期**的基因都需 要表达。然后就要分化了：若**晚期基因**得到表达那么就向裂解途径发展；若阻遏物 C I 蛋白合成被建立，那么就进入溶原化途径。

λ 只有两个早早期基因，*Cro* 和 *N*，由宿主的 RNA 聚合酶转录。

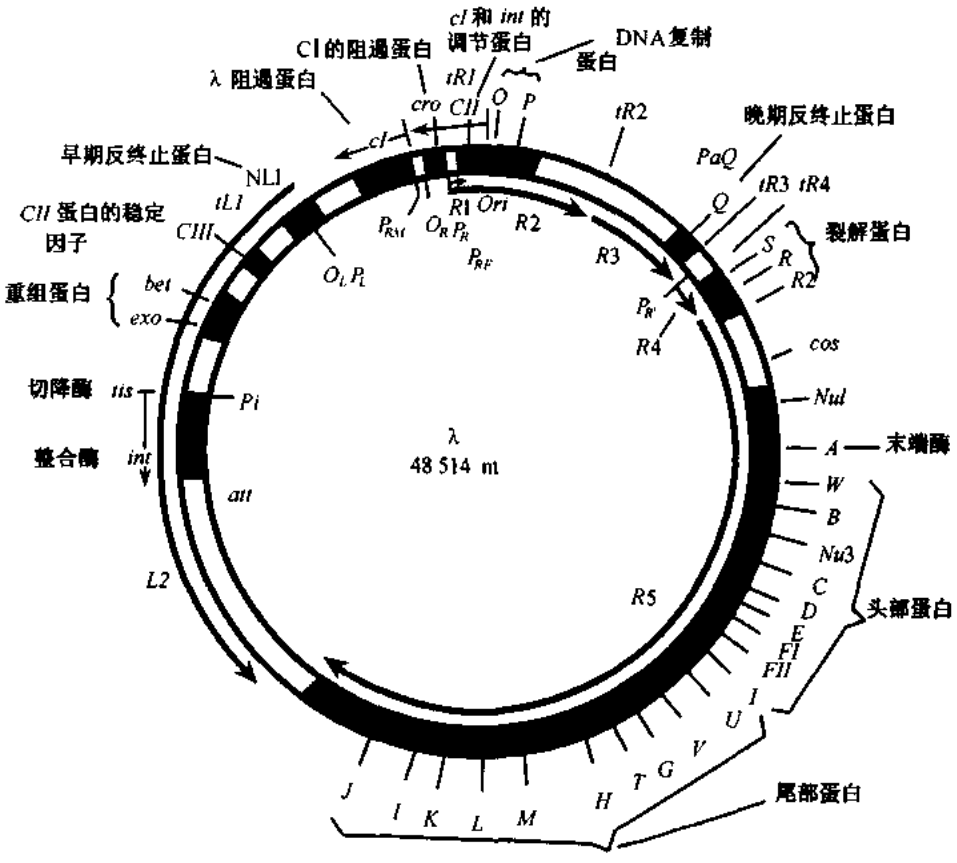


图 17－1　λ 噬菌体的遗传图谱及早晚期的转录。L1.R. 为早期转录物，*R2 L2* 和 *R3* 为晚早期转录物，*R4*, *R5* 为晚期转录物。

N 编码一种反终止蛋白，它作用于 *nut*（N utilization）位点，使早早期转录不会在 t_{R1} 和 t_{L1} 处终止而进成一步转录晚早期的一些基因，其机制我们在后面再系统介绍。

Cro（control of repressor and other things）具有双重功能：它既可阻止 C I 蛋白的合

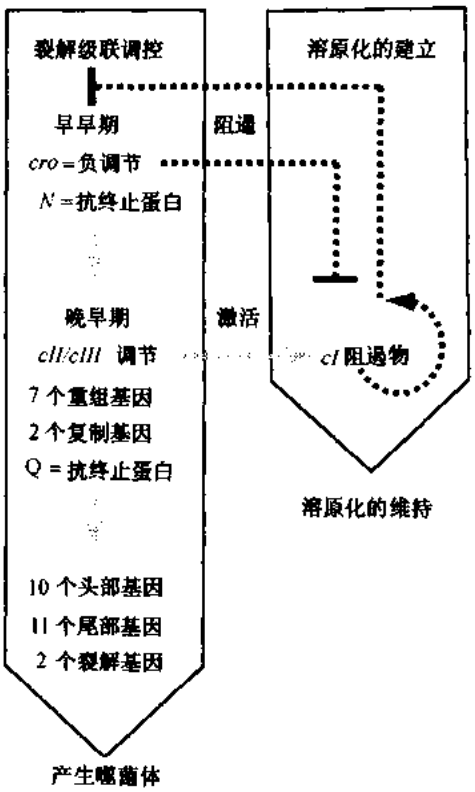


图 17-2 λ噬菌体裂解周期与溶原化周期的联系。(仿 B.Lewin: Genes Ⅵ,1997,Fig.13-7)

成,又可以关闭早期基因的表达(在裂解周期的晚期不需要早期基因的产物)。

晚早期基因包括 2 个复制基因(裂解感染所需的),7 个重组基因(有的和裂解感染中的重组有关,其中有 2 个是 λDNA 整合到细菌染色体中所必需)及 3 个调节基因($c\,Ⅱ$,$c\,Ⅲ$,Q)。它们有着相反的功能。

C Ⅱ和C Ⅲ的复合物可启动 P_E 启动子,对于阻遏物 CI 的开始合成是必要的。而 Q 蛋白也是一个反终止蛋白,它使宿主的 RNA 聚合酶不是在终止子 t_{R3} 处终止,而继续合成晚期基因。因此我们可以将晚期基因分为两类:有的是噬菌体进入溶原化途径所必须的,而另一类是控制进入裂解周期的。

为了搞清楚两种途径,我们首先来考虑裂解途径。图 17-3 提供了 λ 噬菌体 DNA 的图。和调节有关的基因在重组和复制有关的基因之间。在调节基因中,有早早期基因 N 与 *Cro*。它们是由不同的模板继转录的,N 向左,*cro* 向右,N 反终止因子的存在使转录持续向前(穿过终止子)进入重组基因区;向右则进入复制基因区。

图 17-3 是 λ 作为噬菌体颗粒时它的 DNA 组成,但它感染后,DNA 末端相互连接形成了一个环状的 DNA。图 17-4 表明晚期转录中转录了裂解基因等。这些基因在线状 DNA 的右端,而头部和尾部蛋白基因 A.T 在线状 DNA 的右端。

晚期基因是作为单个的转录单位表达的,从位于 Q 和 S 之间的启动子 P_R' 开始,此启动子是连续使用的。但当 Q 蛋白缺乏时,晚期转录终止在 t_{R3} 位点,产生了长 194b

表 17－1　主要的位点及基因的功能

位点或基因	产物及功能
P_L, O_L, P_R, O_R	左右向转录的启动子和操纵子
t_R(1,2,3,4,5)	右向转录的终止子
T_L(1,2)	右向转录的终止子
P_{RE}	C Ⅰ蛋白建立启动子，受 C Ⅱ蛋白调控
P_I	*int*′基因启动子，受 C Ⅱ蛋白调控
P_{aQ}	Q 蛋白反义 RNA 启动子，受 C Ⅱ蛋白调控
P_{RM}	C Ⅰ蛋白基因维持启动子，受 C Ⅰ浓度调控
P_R'	晚期转录的启动子
*nut*L, *nut*R	*N* 蛋白左右两个反终止结合位点
Qut	*Q* 蛋白反终止结合位点
Cro	P_L 和 P_R 的阻遏蛋白，并可阻遏 P_E，抑制 *CI* 表达
c Ⅰ	P_L和 P_R的主要的阻遏物，并可自主调控 P_E
c Ⅱ	可以启动 P_E、P_I和 P_{aQ}，使 λ 进入溶原化途径
c Ⅲ	和 C Ⅱ组成复合物，启动 P_E 产生 C Ⅰ及 *Cro* 的反义 *RNA*
N	t_{R1}, t_{R2}及 t_{L1}的反终止蛋白
Q	t_{R4}的反终止蛋白
O, *P*	*DNA* 复制所需的蛋白
S, *R*	裂解宿主所需的裂解酶
Int	整合酶，使 λ 整合到宿主的染色体中
Xis	切除酶，帮助 λ 在 *att* 位点和宿主连接
bet, *exo*	重组蛋白，帮助 λ 和宿主进行重组
W, *B*, *Nu3*, *C*, *D*, *E*, *F* Ⅰ, *F* Ⅱ, *Z*	头部蛋白基因
U, *V*, *G*, *T*, *H*, *M*, *L*, *K*, *I*, *J*	尾部蛋白基因
Cos(cohesive)	12bp 的回文序列，由线状连接成环状的连接点，A 蛋白切割位点
A	末端酶，识别 *cos* 位点，包装时将环连体切成单个的基因组

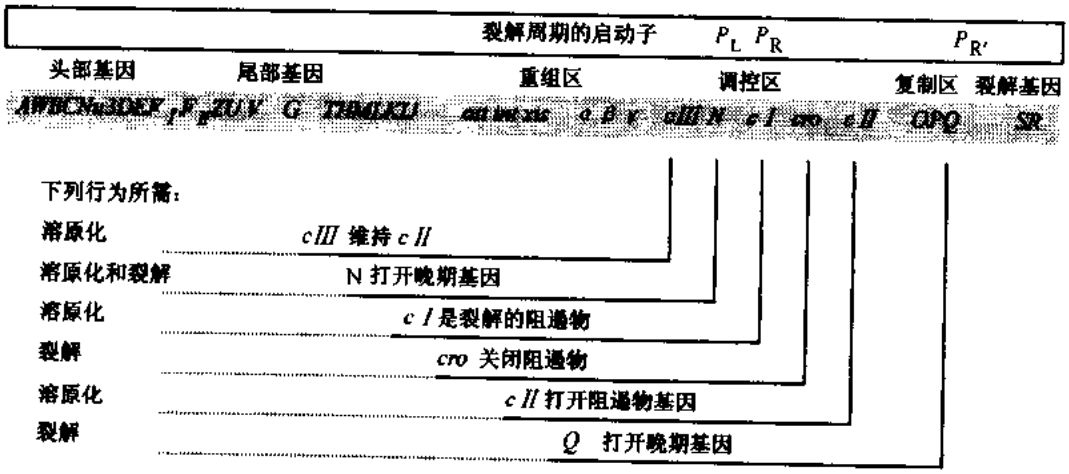


图 17－3　λ噬菌体相关功能的基因簇及在两种途径选择中的作用。(仿 B. Lewin: Genes Ⅵ, 1997, Fig13－8)

图 17-4 λ早期和晚期转录。早早期转录:由宿主的 RNA pol,从 P_L P_R 开始转录 *N* 和 *cro*;晚早期转录:P_R 的反终止作用使转录继续向前在 tR3 终止;晚期转录:P_L 反终止作用使转录持续向前,同时在 P_R'处也起始转录,转录余下的基因。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 13-10)

的转录本,称为 6S RNA。当 Q 蛋白存在时,它阻遏了 t_{R3} 的终止和 6S RNA 的表达,结果晚期基因得到表达。

晚期的反终止的作用与 N 蛋白的反终止作用是相似的。pQ 作用的*qut*(Q-utilization)正好位于晚期启动子的下游,RNA 聚合酶通过此处时,pQ 与其作用消除了终止。(见后)

晚期基因的转录并不在任何特殊位点终止,它转录所有的晚期基因,进入这些基因以外的区域。左边的晚早期转录与此相似,持续穿过重组基因区。每一方向的转录都有可能在聚合酶闯入另一区域时终止。图 17-5 就是比较这两种不同的策略。

二、反终止作用

在控制噬菌体感染的一个共同特点是噬菌体的基因(早期基因)很少是通过宿主的 RNA 聚合酶转录的。但在这些基因中它们的产物作为调节物总是让下一组噬菌体基因得以表达。这种作用的两种普遍的类型是调节蛋白发动新的噬菌体启动子的起始,或是导致宿主的聚合酶通读噬菌体的终止子。

识别新的噬菌体启动子的机制是替换 σ 因子,或者重新合成噬菌体的 RNA 聚合酶(见 17 章)。在这两种情况下,关键的问题是新的一套基因要具有特殊的启动子,它和原来能被宿主 RNA 聚合酶识别的启动子不同。

在启动子识别转换中,特殊的 σ 因子或新的 RNA 聚合酶的合成之后,新的一套基因的表达并不依赖于早期基因的表达,两套转录本是独立的。当发生识别转换时,早期基因的表达能使下一个区域也得到表达(图 17-5)。

终止对于噬菌体控制的从早期基因的表达转换到下一个区域基因的表达提供了一种完全不同的机制。它采用依赖于基因特殊排列的反终止作用。早期基因是与下一个区域的基因相连接的。但二者之间有一个终止子位点。若终止作用在这个位点被阻止的话,那么 RNA 聚合酶就可以通读而进入另一位点。这样在反终止作用中,相同的启动子可以持续被 RNA 聚合酶所识别。因此新的基因是通过 RNA 的延续,形成一个长的 RNA 链而得到表达的。在这种的 RNA 中,5'端是早期基因的序列,3'端是新基因

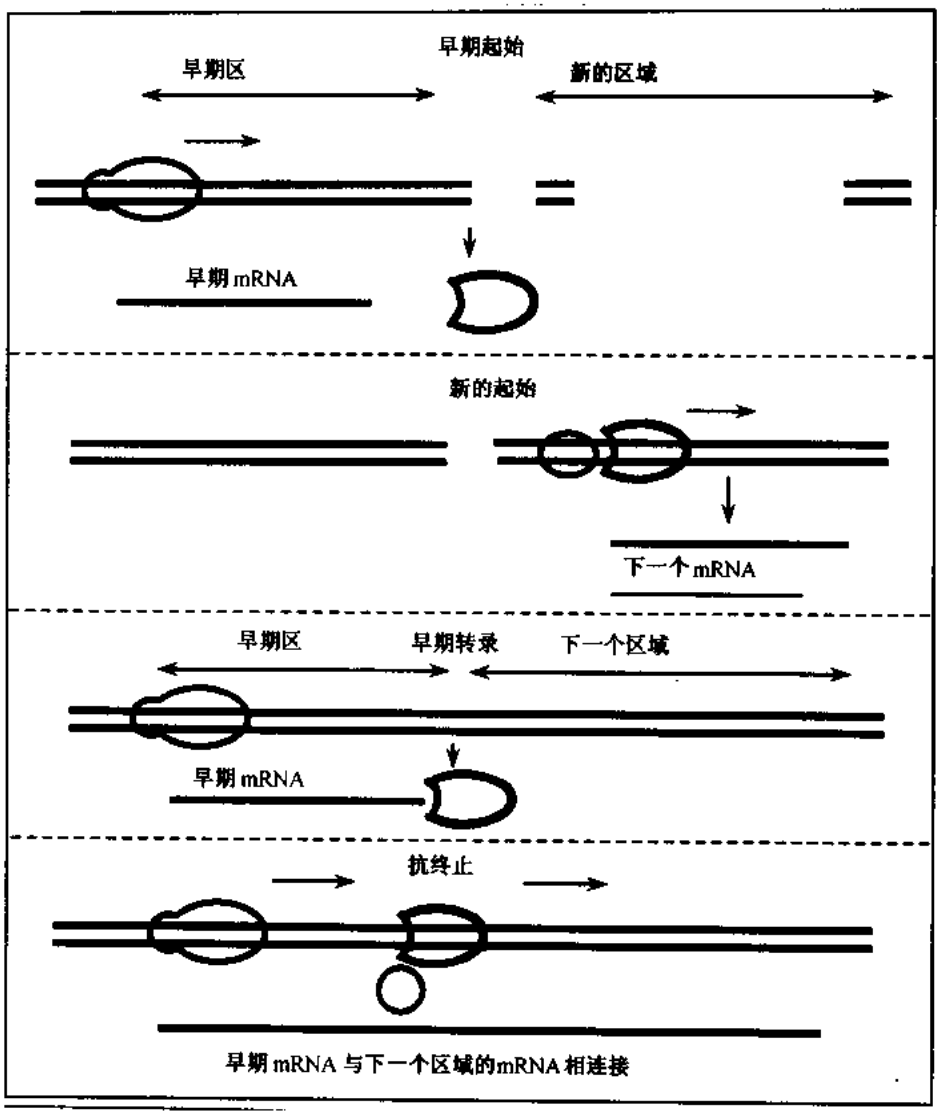


图 17－5　启动子识别转换(上)和抗终止作用(下)两种机制的差别。(仿 B. Lewin: Genes V, 1994, Fig. 16－7)

的序列。由于这两种序列彼此连接在一起,早期基因的表达必然是持续的。

反终止作用最好的例子是噬菌体 λ 提供的。宿主的 RNA 聚合酶起始转录两个**早早期基因**(immediate early genes),下一阶段基因的转录表达是由早期基因末端的终止子控制的。结果**晚早期基因**(delayed early genes)也得到了表达。

控制从早早期基因转换为晚早期基因表达的调节基因是 *N*, 这是通过 *N* 基因的突变鉴别出来的。*N* 基因只能在早早期转录,它的作用不进一步地进入感染循环。看来它与噬菌体 *spoL* 的基因 *28* 的作用相同。从遗传学的观点来看,新起始的机制和抗终止是相似的:两者都是正控制,在调节中一个早期基因的产物的产生是下一套基因表达所必须的。λ 噬菌体早期区的遗传图如图 17－6 所示:*N* 和 *cro* 是早早期基因,它们各自由左向(P_L)和右向(P_R)启动子转录。E. coli 的 RNA 聚合酶进行的转录在 *N* 和 *cro* 基因的末端 T_{L1} 和 T_{R1} 位点分别停下来。接着由于 *N* 基因的表达这种情况发生了改变。*N* 基因的产物 pN 是一种**抗终止蛋白**(antitermination protein)它使 RNA 聚合酶通读

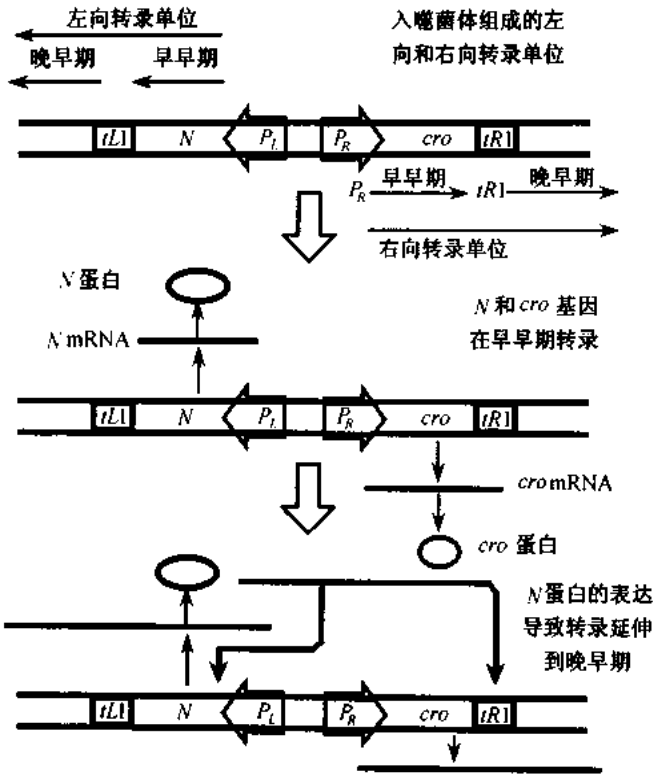


图 17-6 N蛋白的表达。导致转录延伸到晚早期区域。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 13-9)

t_{L1} 和 t_{R1} 而进入了两侧的晚早期基因区。由于 pN 是高度不稳定的,其半衰期仅有 5 分钟,*N* 基因的持续表达是维持晚早期基因的转录所需要的。

就像其他的噬菌体一样,λ 噬菌体晚期基因(编码噬菌体颗粒成分)的表达还需要另外的一些调控。作为晚早期基因之一的 *Q* 基因就可以参与调节。其产物 pQ 是另一种抗终止蛋白,它可以特异地使 RNA 聚合酶起始晚期启动子 P_R',通读它和晚期基因之间的终止子,通过抗终止蛋白的不同特异性可以构成基因表达的级联调节。

pN 和 pQ 的不同特异性建立了一个重要的普遍性原则:RNA 聚合酶与转录单位互相作用通过附加因子能在某些终止子位点特异地发动抗终止作用。终止也是像起始一样以严格的特异性来进行控制。哪些位点和终止的特异性有关呢?

pN 抗终止的活性是高度特异的,但抗终止本身并不取决于终止子 t_{L1} 和 t_{R1};抗终止所需的识别位点位于转录单位的上游,在离终止位点不同的地方。当我们知道某些蛋白在 DNA 的某位点上发挥它们的作用时,我们并不能断定这个位点和起始识别是一致的,它们可以是彼此分离开的。

pN 作用的识别位点叫做 *nut*(*Nut*ilization)。负责左向和左向反终止位点分别是 *nutL* 和 *nutR*。通过 *nut* 突变作图发现 *nutL* 位于 P_L 的起始点和 N 编码区的起始点之间;*nutR* 位于 *cro* 基因末端和 T_{L1} 之间,这意味着两个 *nut* 位点所处的位置离它们转录单位的相对距离是不同的。*nutL* 离启动子此较远,*nutR* 离其终止子比较近。

终止作用是如何发生的呢?pN 识别 *nut* 位点时,它必须对 RNA 聚合酶起作用,使这个酶对于终止子不产生较长时间的反应。*nut* 位置的可变性表明抗终止和起始及终止无关,但通过 *nut* 位点可使 RNA 聚合酶延伸 RNA 链,如图 17-7 所示聚合酶好像变

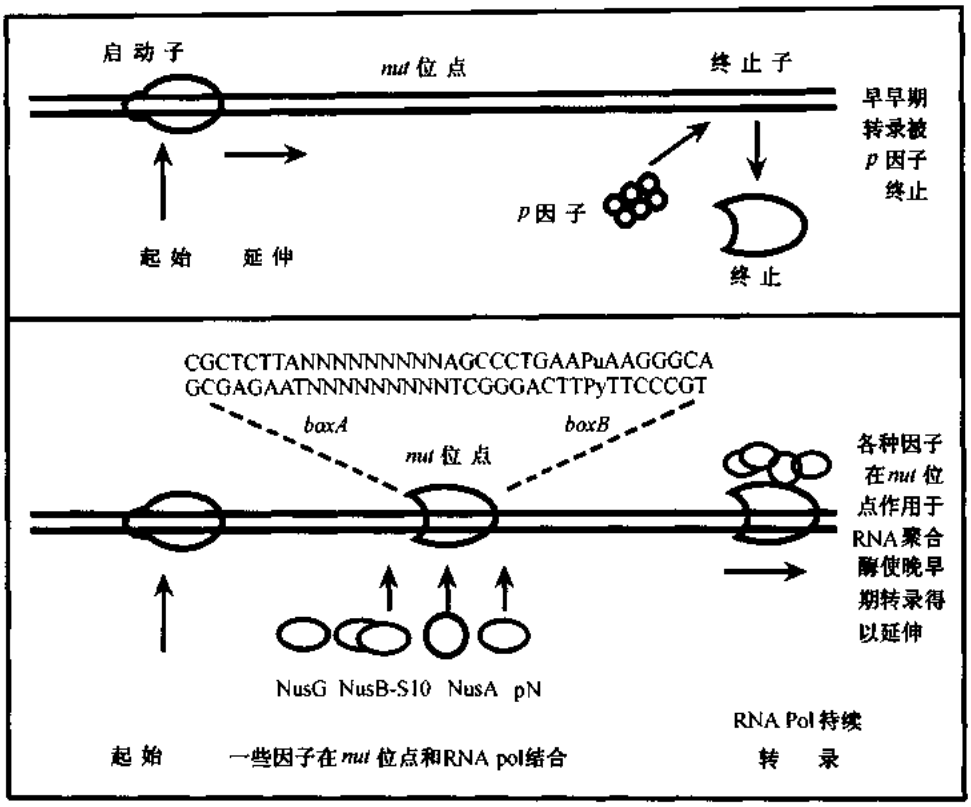


图 17-7　nut 含有两个顺序。核心酶通过 boxA 时，NusB-S10 和它结合，当通过 boxB 时 NusA 和 PN 与 RNA pol 结合。pN 的存在使 RNA pol 通读终止子，产生含有早早期顺序和晚早期顺序和连结 mRNA。（仿 B. Lewin：Genes Ⅵ，1997，Fig. 11-32）

成了穿壁的幽灵，对终止信号毫不理睬而持续穿过。这个反应涉及到在依赖 ρ 因子的终止子上产生抗终止作用，但 pN 也能作用于**内部终止子**(intrinsic terminators)。

pN 识别转录单位中一段短序列的这种能力能否广泛地用于终止机制呢？另一些噬菌体有不同的 *N* 基因和不同的抗终止特异性。*nut* 位点在不同噬菌体中序列也不相同。因此各种噬菌体必须有特殊的 *nut* 位点供自己的 pN 特异识别。各种 pN 产物在终止作用中必须具有与转录装置相互作用的能力。但活化机制的 DNA 序列有不同的特异性。

终止与抗终止是密切相关的，它涉及到细菌和噬菌体的一些蛋白，这些蛋白和 RNA 聚合酶相互作用，对一些转录单位的 DNA 序列作出反应。λ 的 *nut* 位点含有两个序列元件，称为 **A 盒**(*box A*)和 **B 盒**(*box B*)(图 17-7)。A 盒在细菌的操纵子中也有存在，它在噬菌体和细菌的操纵子中对于抗终止是必须的。B 盒的序列发生突变就会消除 pN 的抗终止能力；B 盒只存在于噬菌体基因组中。

抗终止作用作为噬菌体的一种调控机制被发现，导致转录装置其他成分的进一步得到鉴别。细菌中和 pN 相互作用的蛋白质通过分离不受 pN 影响的 *E. coli* 突变体而被鉴别出来。这些细菌的突变株不能被 λ 成功感染。因 λ 感染后仅发生早期基因的表达。这一系列的突变位点都位于 *rpoB* 基因中。此表明 pN 就像 ρ 因子一样，是和核心酶的 β 亚基相作用的。

通过另一种阻止 pN 的功能的 *E. coli* 的突变而鉴别出 *nus* 位点：*nusA*，*nusB*，*nusE*

和 *nusG*。这些 *nus* 位点编码的蛋白形成了转录装置的一部分，但却不能从 RNA 聚合酶中分离出来。*nusA*，*nusB* 和 *nusG* 的功能是单独地和转录终止相关联。*nusE* 编码核糖体蛋白 S10；它在 30S 亚基中的位置与它在终止中的功能并没有明显的关系。

NusA 是一个普通的转录因子，它可增加终止的效率。可能是通过增强 RNA 聚合酶在终止子上停下来的趋势（实际上是在二级结构的另一区域）。NusB 和 S10 形成一个二聚体，特异地和 RNA 的 *boxA* 序列相结合。NusG 可能和所有的 Nus 因子与 RNA 聚合酶聚集成一个复合体有关。Nus 功能的差别是 NusA 足以使 pN 在内部终止子上阻止终止；而在 ρ 依赖性终止子上的抗终止需要所有 4 种 Nus 因子的功能。

抗终止发生在 *E. coli* 的 *rrn*（rRNA）的操纵子中，而且涉及相同的 nus 功能。*rrn* 操纵子的前导序列区域含有 boxA 序列；NusB－S10 二聚体识别这些序列。RNA 聚合酶延伸通过 boxA 时，二聚体与其结合。在此途径中它所通读转录单位中的 ρ 依赖性终止子。

λRNA 的 boxA 序列不能和 NusB－S10 相结合，可能需要 NusA 和 pN 的存在；*boxB* 序列可能是提高反应稳定性所需要的。在 *boxA* 序列中的变异可以被确定，它需要一套特殊的因子进行终止。结果是相同的：RNA 聚合酶通过 *nut* 位点后，它被附加因子修饰，随后它到达终止子时，便可顺利通过，不再终止。

NusA 结合于 RNA 聚合酶的核心酶上，但不结合全酶。当 σ 因子加到 $\alpha_2\beta\beta'$NusA 复合体上时，NusA 蛋白被取代下来，这样又重建 $\alpha_2\beta\beta'\sigma$ 全酶，$\alpha_2\beta\beta'$NusA 准备终止转录，全酶准备重新起始。当全酶结合到启动子上时，它释放出 σ 因子，余下了核心合成 RNA；然后 NusA 蛋白识别核心酶并与其结合，产生 $\alpha_2\beta\beta'$NusA 复合体，当此复合体结合到 DNA 上时，Nus 的成分并不能被取代，但当终止发生时 Nux 因子或被释放、或被 σ 因子所取代。

因此核心酶在起始时与 σ 因子结合，在终止时与 Nus 因子结合，轮流交替。σ 因子和 Nus 因子与核心酶的结合是互不相容的。此二者和核心酶的结合能力可能无明显差异，我们可以把它们看成是轮换的亚基。这样核心酶是 RNA 聚合酶的一种最低量的存在形式，它足以完成 RNA 聚合的基本功能，但缺乏另一些功能。聚合酶在何处终止和开始时转录装置的宽度现在尚不清楚。

抗终止需要 pN 结合到 RNA 聚合酶上，这种结合这在一定程度上取决于转录单位的序列。pN 识别 DNA 上还是 RNA 转录本上的 *boxB* 位点？它并不能直接地和这两种类型的序列结合，但是当核心酶通过 *boxB* 位点时，它和转录复合体上的 *boxB* 结合。很可能它能识别 *boxB* 序列，但为了结合还需要蛋白质和 RNA 聚合酶接触。在参与转录复合体之后，pN 仍结合在核心酶上，实际上它已成为了一种附加的亚基，来改变终止子的识别。

NusA 使 pN 和核心酶之间相连结；pN 的作用是作为一种反终止蛋白阻止 NusA 发挥其功能。N 的突变可以通过 *nusA* 突变而得到恢复；而且这两种蛋白在体外能相互结合。pN 多肽是一种小分子（13.5kDa），碱性，形状是对称的。NusA 和 pN 很快地结合成转录复合体。若没有宿主的蛋白（包括 NusA），pN 不能对 RNA 聚合酶起作用，使其通过 *nut* 位点。在 NusA 存在时，pN 能和 RNA 聚合酶结合。

pQ 可导致的抗终止作用被进一步发现；pQ 是噬菌体调节蛋白，在噬菌体感染的晚期，它阻止转录终止。*qut* 序列是 pQ 作用所需要的，它位于晚期转录单位的起始处。*qut* 的上游部分位于启动中。下游部分位于转录区的起始处，这意味着 pQ 的作用涉及到 DNA 的识别。

pQ 的基本作用是干扰停顿，一旦 pQ 作用在 RNA 聚合酶上，这个酶就明显地在所

有位点减少停顿，包括在 ρ 依赖性终止位点和内部终止位点。所以 pQ 并不直接地作用终止本身，但它可以使酶急速地通过终止子。这样就消除了核心酶或附加因子产生终止的机会。

我们还不清楚在不同操纵子中不同抗终止机制的细节。NusB－S10 能在 *rrn* 操纵子中抗终止，在 λ 中需要 pN。由于这些蛋白质在终止和抗终止功能中作为一种复合体修饰了 RNA 聚合酶的行为，它们个别的反应有时存在着似是而非的情况；NusA 被发现是作为一种 pN 抗终止时所需的因子，但它本身又和终止反应有关。然而一般的原则是 RNA 聚合酶以某种形式存在就足可完成特殊阶段的转录，它们在这一阶段的活性可以只通过修饰而发生改变。σ 因子的更替可以将一种适合起始的形式变成另一种形式。Nus 因子的加盟可能改变了多肽适合终止的形式。

终止看来与延伸模型有着某种神秘的联系。在基本的转录模型中，核心酶在延伸时要经历很多的停顿，而在终止位点的停顿是发生终止的前提条件。在一些因子如 NusA 的干扰下，停顿的时间延长了，从而增加了终止的效率。在 pN 和 pQ 的干扰下，停顿缩短了，减少了终止的效率。由于这些因子的识别位点只存在于一定的操纵子中。停顿和随之发生的终止只会在这些操纵子中发生改变。

第二节　溶原化途径的维持

现在来看一下 λ 的裂解级联调节。一开始 P_L 和 P_R 两个启动子起始转录。早早期基因 *N* 和 *cro* 得到表达。由于抗终止作用，使转录延伸到下一阶段（晚早期基因）的表达。

图 17－8 表示 λ 的调节区，P_L 和 P_R 两个启动子位于 *c I*（clear plaque）基因的两侧。操纵基因 O_L 和 O_R 与 P_L 和 P_R 相联，阻遏物结合于 O_L，O_R 阻遏 RNA 聚合酶起始转录。每个操纵基因的序列都和启动子重叠，以此进行控制。因此常常写为 P_L/O_L 和 P_R/O_R 操纵区。

由于裂解级联的连续性，控制区提供了一个压力点，在这个压力点上能控制是否进入整个的循环。通过拒绝 RNA 聚合酶进入这些启动子，阻遏物阻止了噬菌体基因进入裂解循环。

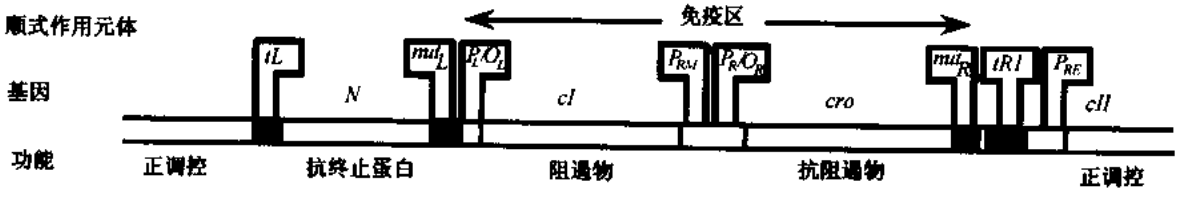


图 17－8　λ 调节区含有一簇反式作用因子和顺式作用元件。（仿 B.Lewin：Genes Ⅵ，1997，Fig. 13－11）

一、C Ⅰ蛋白功能

阻遏蛋白是 *c* Ⅰ基因编码的，这个基因发生突变就不能维持溶原化，而使噬菌体进入裂解循环。这个基因的名字就反应了感染产生的表型。

当细菌被噬菌体感染了，那么细胞被裂解，在平板上产生了一个小而清晰的圆形区域，叫做噬菌斑(plaque)。野生型噬菌体产生的噬菌斑是混浊的或者是模糊的。这只因为它们的一些细胞建立了溶原化而不再裂解。*c* Ⅰ突变的作用是阻止溶原化，这样噬菌斑只含有裂解细胞，因此感染后只产生清晰的噬菌斑(Clear plaque)。有 3 个基因(*c Ⅰ*、*c Ⅱ*、*c Ⅲ*)的名字和此种表型有关。

c Ⅰ 基因是以位于其右端的启动子 P_{RM} 开始转录的(RM 意思是阻遏物的维持，repressormaintenmce)，转录的终止是在基因的左端。其 mRNA 实际是从 AUG 起始密码子开始；由于缺乏通常的核糖体结合位点，此 RNA 的翻译效率不高，仅产生低水平的阻遏蛋白。

C Ⅰ蛋白独立地结合于两个操纵子上。结合在 O_L 上可阻止 RNA 聚合酶与 P_L 的结合；结合在 O_R 上不仅阻止 RNA 聚合酶结合在 P_R 上，同时促进了 RNA pol 结合于 P_{RM} 上。

CI 阻止 RNA 聚合酶在 P_L 起始转录，从而停止了 *N* 基因的表达。由于 P_L 是用于所有左向早期基因的转录，因此此作用阻止了左向早期转录单位的表达。这样在早期阶段的基因表达前裂解周期就已被阻断了。

阻遏蛋白结合在 O_R 上阻碍了 P_R 的使用。这样 *cro* 和其他的右向早期基因也不能表达。这是很重要的，因为当溶原化得以维持时可阻止 *cro* 的表达。当阻遏物结合在 O_R 时，RNA 聚合酶才能起始 P_{RM}。此时的阻遏物的行为是作为正调节蛋白，对 *cI* 基因的转录是必要的。由于阻遏物是 *c Ⅰ* 的产物，这种相互作用产生了正的自主调节循环。在此循环中，阻遏物的存在对它本身的持续合成是必须的。

这种控制循环的特点是解释了溶原化的生物学性质。溶原现象是稳定的，由于控制循环使得只要阻遏物的水平达到一定的量，*c Ⅰ* 基因就能持续表达。结果是 O_L 和 O_R 仍被阻遏物持续地占据着。进入裂解极联被阻滞了。λ 维持了原噬菌体状态。

阻遏的存在解释了免疫的现象。所谓 λ 的免疫是当噬菌体 DNA 再次进入溶原化细胞时，细胞中原有的原噬菌体基因组合成的阻遏蛋白免疫性地结合在新基因组的 O_L 和 O_R 上，这次就阻止了第二个噬菌体进入裂解周期。

此操纵基因最初通过**烈性**(virulent)突变被鉴定为是阻遏物的靶位点。这种突变阻止阻遏物结合到 O_L 和 O_R 上，结果当噬菌体感染新的宿主时，噬菌体必然就会进入裂解途径。λ *uir* 突变能在**溶素原**(lyxogens)上生长，因为在 O_L 和 O_R 中的烈性突变使产生的噬菌体不理会残留的阻遏物而进入裂解周期。噬菌体的烈性突变是在细菌操纵子中操纵基因的结构突变。

原噬菌体当溶原化途径被阻断时被诱导进入**裂解周期**。当阻遏物失活时，也会发生此现象。在缺乏阻遏蛋白时，RNA 聚合酶结合到 P_L 和 P_R 上，开始了裂解循环(图 17－9)。

阻遏物维持的自主调控产生了一个敏感的反应。由于阻遏物的存在对其本身的合成是必须的，存在的阻遏蛋白一旦破坏，*c Ⅰ* 基因的表达就立即停止。这样阻遏物不能合成，其分子就会被降解。所以没有溶原化的干扰，裂解周期就可以开始。

c Ⅰ 基因和 *cro* 基因的左、右操纵基因区域决定着噬菌体的免疫。具有这一区域的

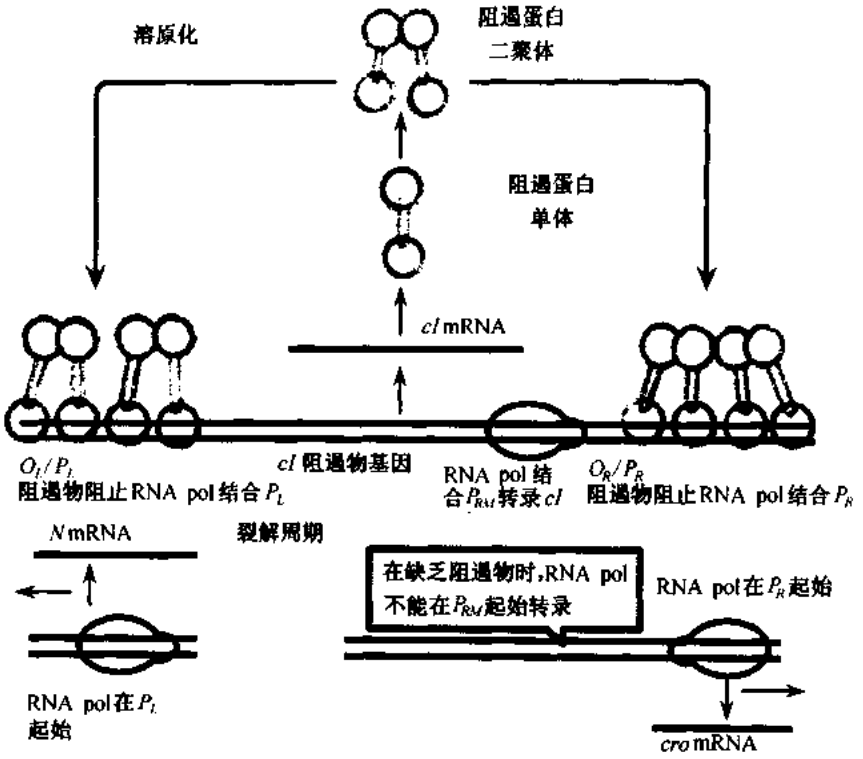


图 17－9　通过自我调节维持溶原化(上);若此循环中断,裂解周期就开始了(下)。
(仿 B.Lewin:Genes Ⅵ,1997, Fig.13－13)

任何噬菌体都有相似类型的免疫。这是由于阻遏蛋白及其作用位点二者都是特异的。因此这一区域就称为**免疫区**(immuinty region)(图 17－8)。以希腊字母命名的噬菌体 φ80、φ21、φ434 和 λ 各自都只有一个免疫区。我们说溶原化的噬菌体对任何其他相同类型的噬菌体是免疫的,更准确切的意思是:免疫是对任何其他具有相同免疫区的噬菌体而言。

二、CⅠ蛋白的结构

CⅠ蛋白的亚基是一个分子量为 27kDa 的多肽,具有两个特殊的功能区(图 17－10)。① N 端功能区,从第 1～92 氨基酸,提供了操纵基因结合区。②C 端功能区,132～236

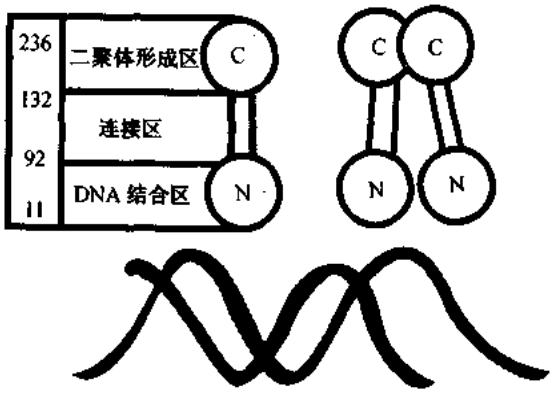


图 17－10　CI 蛋白的结构。(仿 B.Lewin:Genes Ⅵ,1997,Fig.13－14)

氨基酸，负责形成二聚体。

这两个功能区由长氨基酸(从 92～132 氨基酸)的连接区相连接。当阻遏物被蛋白酶降解时，每个功能区成了彼此分离的片段。二个功能区都能独立发挥功能。C 末端片段能形成寡聚体。N 末端片段能与操纵基因结合。但亲和力要比完整的阻遏物低。这样与 DNA 特异性结合的信息是在 N 端功能区，但通过 C 端功能区的附着可以增强与 DNA 的结合效率。

阻遏物的二聚体结构对于维持溶原化来说是极为重要的。溶原化的原噬菌体经诱导进入裂解周期是由于阻遏物的连接区(在 111 到 113 氨基酸之间)被切断所致。这是构象改变的一种类型。

完整的 CI 蛋 C 端功能区形成二聚体，当它结合到 DNA 上时两个 N 端功能区同时与DNA 接触。但被剪切的 CI 蛋白失去了 C 端，N 端不能形成二聚体(图 17－11)打乱了单体和二聚体之间的平衡，这样阻遏物就从 DNA 上解离下来，随着就进入了裂解周期。

在溶原化和裂解周期之间的平衡取决阻遏物的浓度。在溶原化的细胞中完整的阻遏物其浓度足以占据操纵基因。但如果阻遏物被剪切的话，这个浓度就不够了。因为分离的 N 端功能区对操纵基因的亲和力是较低的。阻遏物的浓度如果较高的话是不会因剪切而诱导裂解周期的。

阻遏作用依赖于 CI 蛋白的浓度，因此阻遏物的行为对阻遏作用的干扰很大。λ 的阻遏物仅以二聚体的形式和 DNA 结合。单体和双体之间的平衡取决于蛋白质的浓度。平衡的结果是 CI 蛋白在低浓度时以单体存在；在高浓度时是以二聚存在。结果 CI

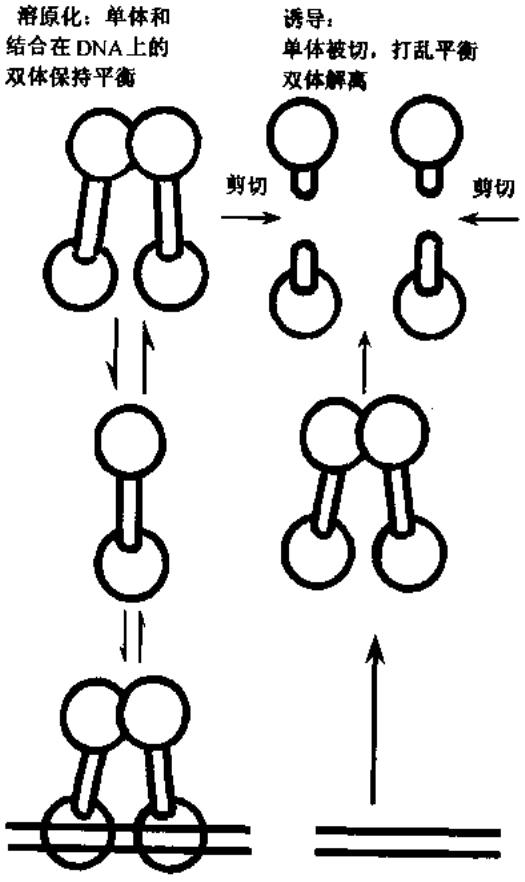


图 17－11　溶原化和诱导状态的阻遏物。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 13－15)

蛋白对操纵基因的结合取决于单体浓度的平方。

λ 噬菌体阻遏物浓度的减少使平衡向单体方向移动，这样就不能有效地与 DNA 结合，与操纵子结合能力的下降。图 17－12 表示阻遏物－操纵基因结合曲线，呈典型的二级反应，对阻遏物浓度的反应要比一级反应更为敏感(在一级反应中，阻遏物的结构不变)。通常在溶原状态下 λ 噬菌体阻遏物的浓度是约 200 分子/细胞，相当于 4×10^{-7}mol/L 左右。无论在什么时候，这个浓度都足以完成阻遏作用。若浓度下降 10 倍左右，将有效地解除阻遏，而产生诱导。

三、阻遏物与 DNA 的结合及调控机制

各种调节细菌转录的 DNA 结合蛋白都显示了和 DNA 结合的相似模型。在此模型

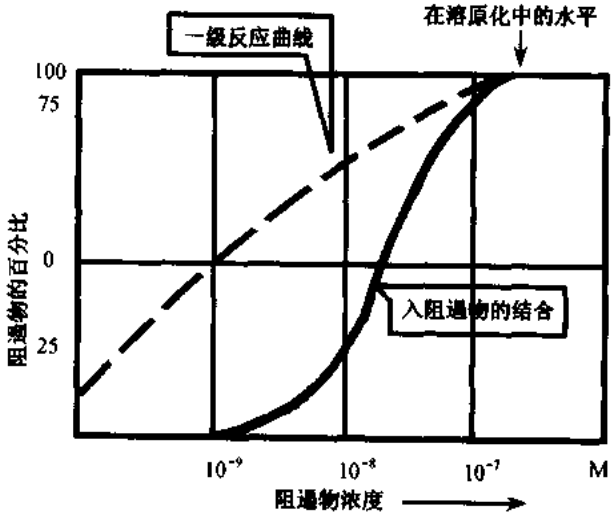


图 17－12　λ 阻遏蛋白与操纵位点结合的二级动力学。
(仿 B.Lewin:Genes Ⅵ,1997, Fig.13－16)

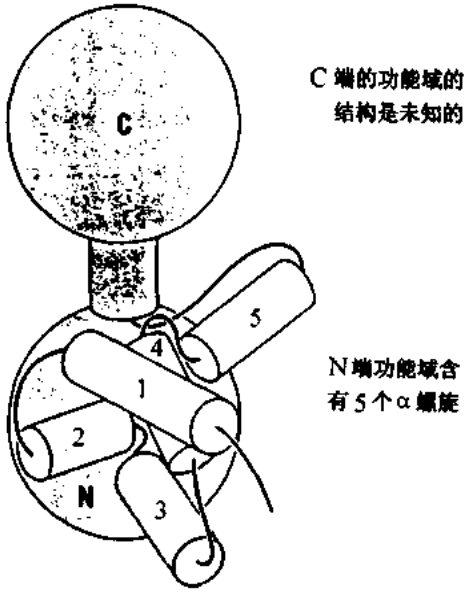


图 17－13　λ 噬菌体阻遏蛋白的 N 端功能域含有 5 个 α 螺旋，α 螺旋 2 和 3 涉及和 DNA 的结合。
(自引 B.Lewin: Genes Ⅵ,1997,Fig.13－17)

中活化的功能域含有 2 个短的 α 螺旋区与 DNA 接触。(在真核细胞中有些转录因子使用相同的基序)。

λ 噬菌体阻遏蛋白的 N 端功能区含有几个 α 螺旋,其排列如图 17－13 所示,它的连接区和 C 端的功能区的结构尚不清楚。

N 端的 2 个 α 螺旋区负责和 DNA 结合。DNA 接触是以**螺旋转角螺旋**模型(图 17－14)进行的。注意单个的单体,α－螺旋－3 含有 9 个氨基酸,它和前面 α－螺旋－2(含 7 个氨基酸)呈一定的角度。在二聚体中两个并列排列的螺旋－3 区域相距 34Å。这样适合于他们嵌入相邻的两个 DNA 大沟中。在这两个区 α－螺旋基序的相关排列在多种 DNA 结合蛋白中都有所发现。包括 CAP,lac 阻遏物和各种噬菌体的阻遏蛋白。每个 α 螺旋在与操纵基因的结合中都起到独特的作用。

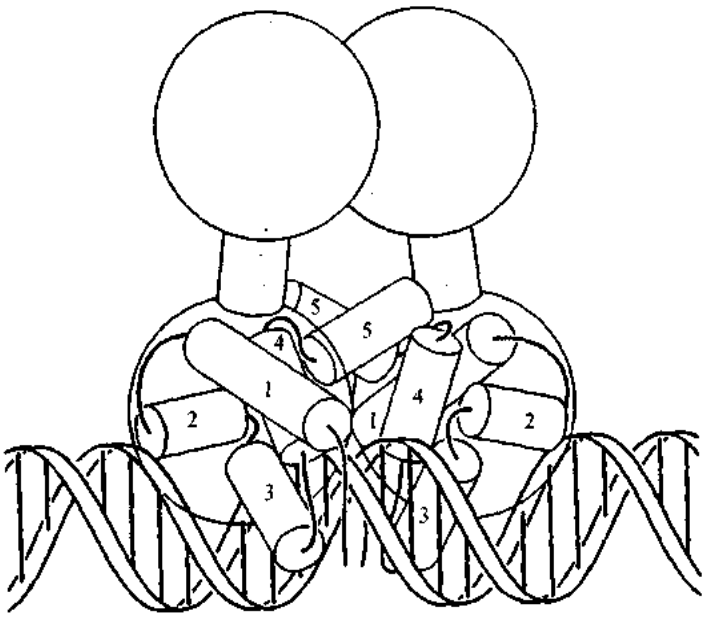


图 17－14　在二聚体的 α－螺旋和 DNA 结合的模型中,α－螺旋－3 单体位于 DNA 同一面的大沟中,而 α－螺旋－2 横在此大沟上。(引自 B. Lewin: Genes Ⅳ, 1990, Fig. 16－16)

在螺旋－3 和 DNA 的相互作用中要求于氨基酸链和暴露的碱基对之间的氢键。这个螺旋负责识别特殊的 DNA 靶序列。因此也称它为**识别螺旋**(recognition helix)。

螺旋－2 和 DNA 的接触产生的氢键和磷酸骨架相连。这些相互作用对于结合是必要的。但并不控制靶识别的特异性。除了这些连接之外,和 DNA 相互作用的大部分能量是由离子与磷酸骨架相互作用所提供的。

如果改变此编码区的序列,通过替换构建一个新的蛋白,在这个阻遏物中的识别螺旋的行为将如何呢? 这种杂种蛋白是一种新的识别螺旋。这一短区域的氨基酸序列决定了单个蛋白的序列特异性。且能和其他的多肽链起连接作用。

图 17－15,表明两种蛋白结合相似 DNA 序列的细节。两种 λ 阻遏蛋白 C Ⅰ和 Cro 蛋白有相似的螺旋转角螺旋模体。但它们各自对 DNA 的特异性尚未鉴别。

每种蛋白都具有疏水氨基酸之间类似的相互作用来维持螺旋 2 和螺旋 3 的相互关系。C Ⅰ有一个 Ala－Val 连接,而 Cro 有一个 Ala－Ile 连接(图 17－16)。

在 C Ⅰ螺旋－3 的氨基酸与操纵拉点上的特殊碱基接触,C Ⅰ的 3 个氨基酸识别 DNA 上的 3 个碱基。在 Cro 蛋白中有 5 个氨基酸识别 DNA 上 5 个(有可能是 6 个)碱基。

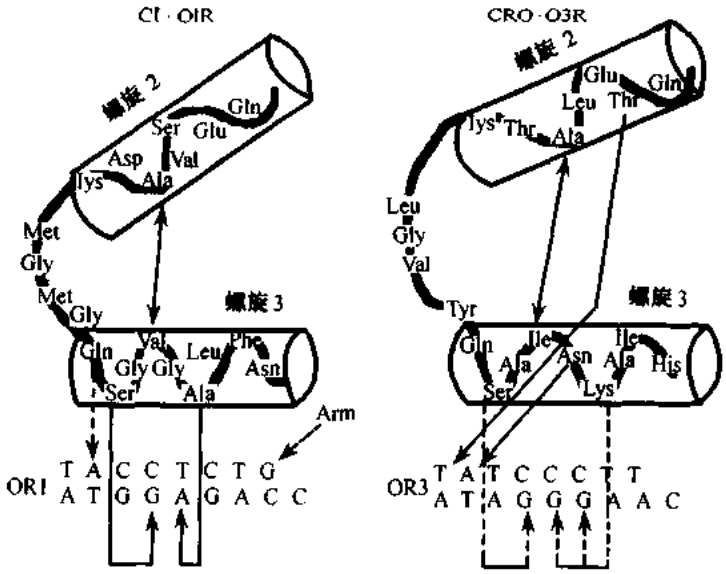


图 17－15　CI 和 Cro 蛋白用 2 个 α 螺旋来和 DNA 接触，识别人操纵子其亲和性是由螺旋 3 的氨基酸序列所决定。(仿 B.Lewin:Genes Ⅵ, 1997, Fig. 13－19)

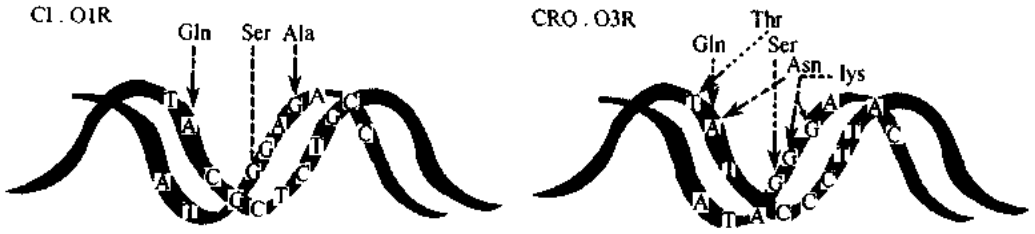


图 17－16　C Ⅰ 和 Cro 螺旋 3 和 DNA 的一个面接触。

在 C Ⅰ和 Cro 中有 2 个氨基酸（螺旋末端的 Gln 和 Ser）与识别有关，而另一些和 DNA 接触的氨基酸是不同的(在 C Ⅰ是 Ala，而在 Cro 是 Asn 和 lys)，在 Cro 中的螺旋 2 中 Thr 也直接和 DNA 接触。

图中表明每种蛋白对 DNA 的识别，结合都是非常严格的；C Ⅰ与 Cro 二者有类似的氨基酸残基，但和 DNA 上不同碱基的连接，表明相关的识别螺旋指导相关 DNA 序列的识别。这就解释了为什么阻遏蛋白和 Cro 蛋白二者识别相同操纵基因的碱基，但对特异操纵基因来说它们具有不同的亲和性。这些序列(图 17－18)在 9 个碱基对中 3 个不同。

被 C Ⅰ或 Cro 蛋白螺旋－3 所接触的碱基都是位于 DNA 的一个表面，这点从图 17－15 中可以看出。但阻遏蛋白还有一个接触点在 DNA 的另一表面上。被分离阻遏蛋白的 N 端功能区与 DNA 的接触和完态 C Ⅰ蛋白相同。若切除掉 N 端最后边的 6 个氨基酸(螺旋 1 突出来的部分)就消除掉某些和 DNA 的接触，从这一实验结果表明，N 末端的功能区的大部分是和 DNA 的一个表面接触的，而 N 端最后面的 6 个氨基酸形成一个手臂伸到 DNA 的背面(图 17－16，17－17)。臂上的 Lys 与大沟中的 G 接触。也和磷酸骨架接触。臂与 DNA 之间的相互作用对 DNA 的结合起了很大的作用。阻遏物若失去了此臂，那么亲和力将下降 1000 倍左右。

每个操纵基因都含有 3 个阻遏物结合区。如图 17－18 所示，每个结合区是由 17个

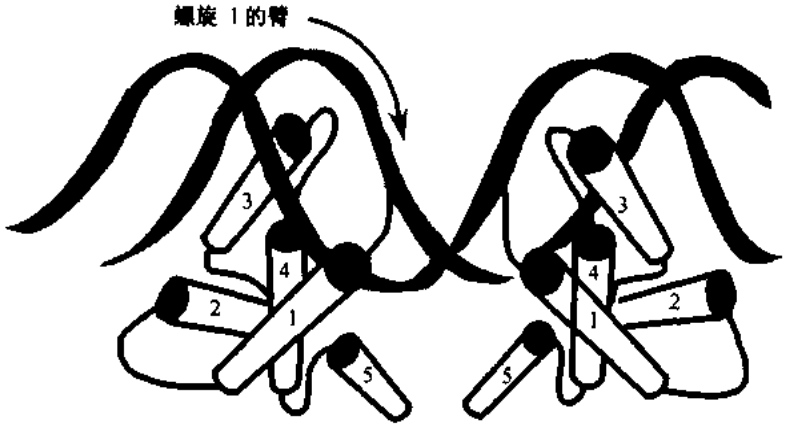


图 17－17　阻遏蛋白螺旋 1 的末端延伸部分与 DNA 的背面接触。
（仿 B.Lewin:Genes Ⅵ,1997, Fig.13－20）

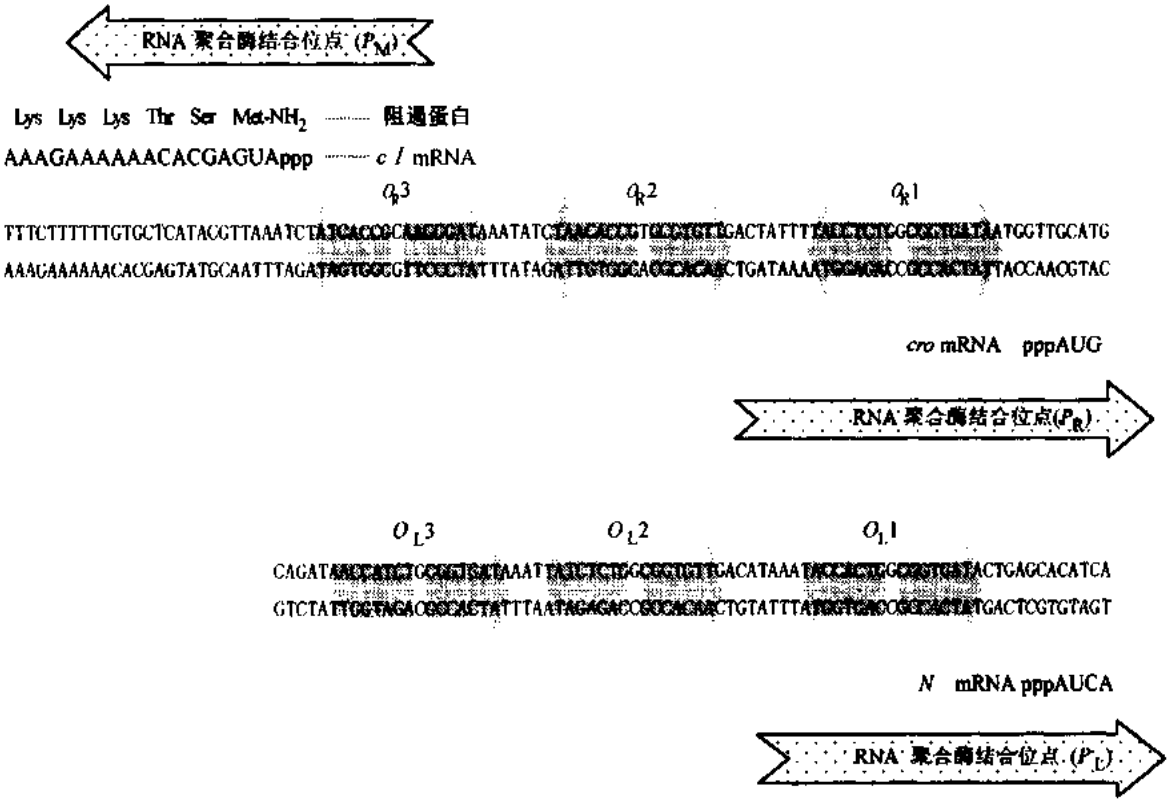


图 17－18　每个操纵基因含有 3 个阻遏物结合位点，而且在 RNA 聚合酶结合位点与启动子重叠。
（仿 B.Lewin:Genes Ⅵ,1997,Fig.13－21）

碱基组成的旋转对称序列。在每个操纵基因中的结合区被 3～7bp 的 A-T 丰富区所分隔，O_R 位点含有 $O_{R1}-O_{R2}-O_{R3}$，O_L 位上含有 $O_{L1}-O_{L2}-O_{L3}$。二者的 1 区都紧靠着启动子的转录起始点，而 2 和 3 区离上游远一点。

阻遏物二聚体对称地结合在回文区域上，每个二聚体的 N 端功能区都和相同的碱基位点相接触。因此单个的 N 端区和半个结合区相接触。这些碱基并不直接和阻遏物蛋白相接触，但对结合有重要的影响。

噬菌体 φ434 的阻遏物以螺旋－转角－螺旋的方式和 DNA 结合，结晶结构表明螺旋－3 被定位在每半个结合区上。因此它和最外测的 5 个碱基接触。而不和里面的 2个

碱基接触。然而，操纵基因在内侧带有 A-T 丰富区，φ434 阻遏物与它结合的能力比与带有 G.C 丰富区的操纵基因要强得多。这是由于 φ434 阻遏物在操纵基因的中心与扭曲 DNA 的结合是很弱的。DNA 的前半个区和后半个区域之间的角度增加了 3°。这可能使阻遏物二聚体的每个单体与 DNA 最适结合。A-T 碱基对使这扭曲比 G-C 对快得多，这就影响了操纵基因对阻遏物的亲和力。

在每个阻遏中的螺旋 3 的序列与它识别的操纵基因的相关碱基序列有密码存在吗？现在这个概念已被抛弃。λ 和 φ434 阻遏蛋白或 Cro 蛋白进行 DNA 结晶结构分析提示了这种作用实际上是和此模型是不相吻合的。在蛋白质中有多个氨基酸和单个碱基产生接触：在蛋白结合后构象随之发生改变。蛋白质和 DNA 的接触模式一般却是相似的。但绝不是丝毫不差。蛋白质特殊位点的氨基酸与 DNA 中的特殊位点接触的模式并非是简单的。

在每个操纵基因上都具有三个重复的接合区，阻遏蛋白开始结合时选择哪一个结合区呢？每个操纵基因的 1 区比其余的结合区具有更高的亲和性（约 10 倍），因此阻遏物总是首先和 O_{L1} 和 O_{R1} 结合。

在每个操纵基因中 λ 的阻遏物结合其余两个区域时都具有**协同效应**。当 1 区存在二聚体时它增加了第 2 个二聚体和 2 区的亲和力。当 1 区和 2 区都被占据时，对 3 区并不起作用。在溶原化时，阻遏物通常都结合在操纵基因 1 区和 2 区，并不占据 3 区。若 1 区基因突变而失活，那么阻遏蛋白就协同地结合于 2 区和 3 区。也就是说 2 区的结合可以帮助另一个二聚体结合到 3 区上。这种相互作用是直接地发生在阻遏物二聚体之间，而并不涉及到 DNA 构象的改变，有可能第一个阻遏物的连接区使二聚体的 C 端朝向第二个二聚体的 C 端。

协同结合的结果使阻遏物在生理浓度下增加了和操纵基因的亲和力。这样使低浓度的阻遏物也能占据操纵基因。这对于在系统中解除阻遏具有不可逆的结果是重要的。在一个操纵子上编码了代谢酶，若阻遏消除就会合成不必要的酶。但 λ 原噬菌体若阻遏失败，将导致噬菌体被诱导使细胞裂解。从图 17－18 中我们可以看到 O_{L1} 和 O_{L2} 大致位于 P_L 和 P_R 中 RNA 聚合酶结合位点的中心。O_{L1} 和 O_{L2} 和 O_{R1}－O_{R2} 被占据，就全阻断 RNA 聚合酶移向启动子的通路。

对 CI 的转录而言启动子 P_{RM} 和 O_R 之间的位置关系和上述的情况不同，RNA 聚合酶的结合位点正好和 O_{R2} 相邻接。这就解释了阻遏物如何自主调节它本身的合成。当两个二聚体结合在 O_{R1}－O_{R2} 时，在 O_{R2} 上的二聚体和 RNA 聚体酶相互作用（图 17－19）。这种作用并不像阻遏物之间的相互作用，而是依赖于 N－端的功能域促进 P_{RM} 的启动。O_{R2} 被占据时此功能域可能作为一个独立的片段。

通过突变消除正控制来对 CI 基因作图。一个有趣的突变类型是仍能保持和操纵基因结合来阻遏转录，但不能促进 RNA 聚合酶从 P_{RM} 转录。此突变发生在一小组氨基酸中，位于螺旋－2 的外侧或者说在螺旋 2～3 之间。此突变使这个区域的负电荷减少。相反，增加负电荷的突变能加强 RNA 聚合酶的活性。这表明这一组氨基酸构成了“酶性补丁”，其功能是 和 RNA 聚合酶上的碱性区域相互产生静电作用。

图 17－19 表示阻遏蛋白中的正控制突变，它们在阻遏物上紧靠 DNA 的磷酸基团，此基因又紧挨着 RNA 聚合酶。在阻遏物上涉及到正调控的这组氨基酸与 RNA 聚合酶接触。阻遏物与 RNA 聚合酶之间的相互作用，对于 RNA 聚合物产生的闭合复合体转变成开放复合体来说是必须的（表 17－2）。

正调控有关的蛋白质之间的相互作用可能涉及到蛋白质各个区域。

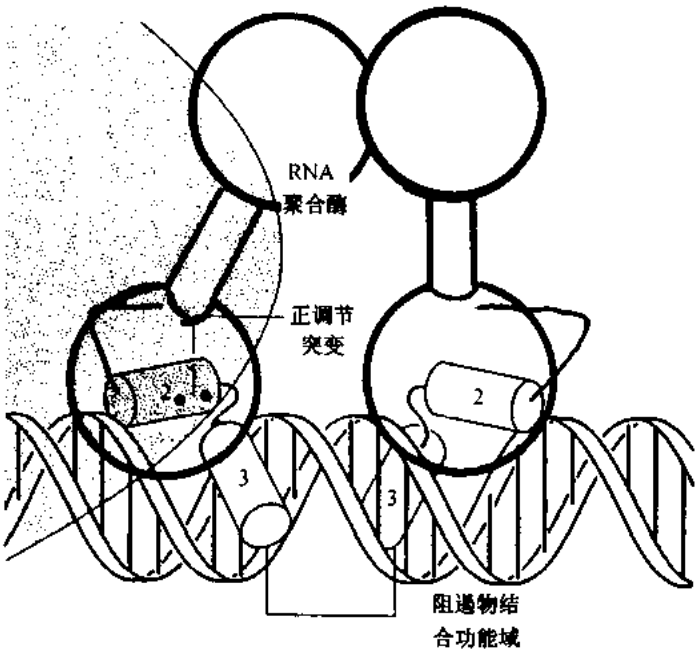


图 17-19 在与 RNA 聚合酶直接作用的 α-螺旋-2 的一个小区域中发现了正调节突变。
(引自 B. Lewin: Genes Ⅳ, 1990, Fig. 16-21)

表 17-2 正调控可在转录启动的任一阶段影响 RNA 聚合酶位点

位点	调节物	聚合酶的结合(平衡系数, K_B)	闭合-开放转换(速率系数, K_2)
λP_{RM}	CⅠ	不影响	11×
λP_{RM}	CⅡ	100×	100×

在噬菌体 P_{22} 中,和正调控有关的突变位于螺旋-3。P_{22} 阻遏物的 DNA 结合结构域与 λ 中是十分相似的,一个共同的原则是蛋白质之间的相互作用可以释放能量用来帮助起始转录。

若一个阻遏物二聚体结合到 O_{R3} 将会引起什么后果呢? 这一区域和 RNA 聚体酶在 P_{RM} 上的结合位点重叠。若阻遏物的浓度大到足以占据 O_{R3} 的活。CI 的转录将被阻断。导致在适当的时候减少阻遏蛋白的浓度。O_{R3} 又被空出来,由于 O_{R2} 仍被占据,所以自我调节的循环又可以重新开始。要求阻遏物有较高的浓度,这一机制又使阻遏物的浓度不至太高。因此阻遏物是表达特殊的一种自我调节蛋白,在低浓度时它的功能是正调控,而在高浓度时它的功能是起负调控作用 。

在 O_L 和 O_R 的 1 区及 2 区发现有烈性突变(可使宿主裂解)位点。这种烈性程度很高的突变是由于阻遏物结合位点亲和性减低之故。但也取决于作用位点和启动子的关系。在 O_{R3} 和 O_{L3} 上通常不会发生烈性突变的。

第三节 阻遏物合成的建立

对于溶原化的维持调节似乎存在着一种自相矛盾,阻遏蛋白的合成是需要阻遏蛋白的存在。这表明溶原化的条件是一直存在的。但阻遏蛋白最初如何建立它的合成呢?

当一个 λDNA 进入新的宿主细胞时，RNA 聚合酶并不能转录 *c Ⅰ*，因为没有阻遏物结合在 P_{RM} 上，而 P_R 和 P_L 此时也没有阻遏物的结合。因此 λ 感染后首先转录的是 *N* 和 *cro*，然后 N 蛋白使转录进一步延伸，当 *c Ⅱ* 和其他基因在右边被转录时，*C Ⅲ* 基因也在左边被转录(图 17－20 所示)。

c Ⅱ 和 *c Ⅲ* 基因共同促进 *PRE* 的启动。如果它们发生突变的话，就会出现清晰的噬菌斑，意味着产生裂解。所不同的是 *c Ⅰ* 的突变不仅不能建立溶原，而且也不能维持溶原。而 *c Ⅱ* 和 *c Ⅲ* 的突变仅对建立溶原化有影响，而一旦建立，就可以通过 *c Ⅰ* 自我调节来维持，无需求助于 *c Ⅱ* 和 *c Ⅲ* 产物。

这表明 *c Ⅱ* 和 *c Ⅲ* 基是正调控基因。阻遏蛋白的两个合成系统中的一种是需要它们的产物。这个系统仅在 *cI* 表达起始时才需要，以此弥补其自我调节不能从头合成的特点。

C Ⅱ蛋白直接作用于基因表达。在 *cro* 和 *c Ⅱ* 基因之间有一个启动子叫做 P_{RE}，右下角"$_{RE}$"的意思是阻遏建立(repressor establishment)。这个启动子只有在 C Ⅱ蛋白存在时才能被 RNA 聚合酶所识别。其作用如图 17－20 所示。

C Ⅱ蛋白在体内是极不稳定的，当它受到一种宿主蛋白 HflA 的作用时就会降解，C Ⅲ蛋白的作用是保护 C Ⅱ免受降解。

从启动子 P_{RE} 的转录建立溶原化是以两种途径进行的。一种是直接作用，能转录

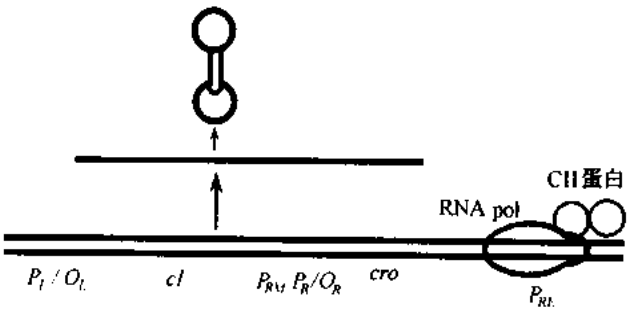


图 17－20　CII 和 RNA pol 起始 P_{RE}，合成 CI 蛋白。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 13－23)

cI 基因，产生阻遏蛋白。另一种是间接作用，转录以反方向通过 *cro* 基因，RNA 的 5′端部分与 *cro* 的转录本相应部分是反义的，与 *cro* 的 RNA 互补结合，抑制其翻译。这样通过反义 RNA 的作用对 *cro* 的表达起到负调节的作用。而 Cro 蛋白又是裂解周期所必须。结果是阻遏裂解周期，有利于溶原化周期的建立。

在 P_{RE} 转录本上的 *cI* 编码区转译的效率要比 P_{RM} 转录本上的高得多。其实 P_{RE} 转录本阻遏物的合成要比 P_{RM} 转录本高约 7～8 倍。这反映了 P_{RE} 转录本具有一种高效率的核糖体结合位点，而 P_{RM} 转录本却没有核糖体结合位点，而是用 AUG 起始密码子来启动的。

P_{RE} 启动子有一个不太好的－10 序列，而缺乏－35 序列。此还不足以解释它对 C Ⅱ的依赖。这个启动子在体外不能由 RNA 聚合酶转录。但加入 C Ⅱ时，就能转录，C Ⅱ的结合区域约从－25 到－45。当加入 RNA 聚合酶时，从－12 到＋13 这一个较大的区域受到了保护。实际上这两种蛋白的结合位点是重叠的(图 17－21)。

虽然启动子中缺乏保守序列，但通过 *cy* 突变的存在表明了此－35 和－10 区域对启动子的功能是重要的。这些突变为两组：*cyL* 和 *cyR*。*c Ⅱ* 和 *c Ⅲ* 的突变同样能起到阻断溶原化建立的功能。但它们以顺式作用代替反式作用。

cyL 突变位于－10 附近，它可能会阻断 RNA 聚体酶识别启动子。*cyR* 突变位于

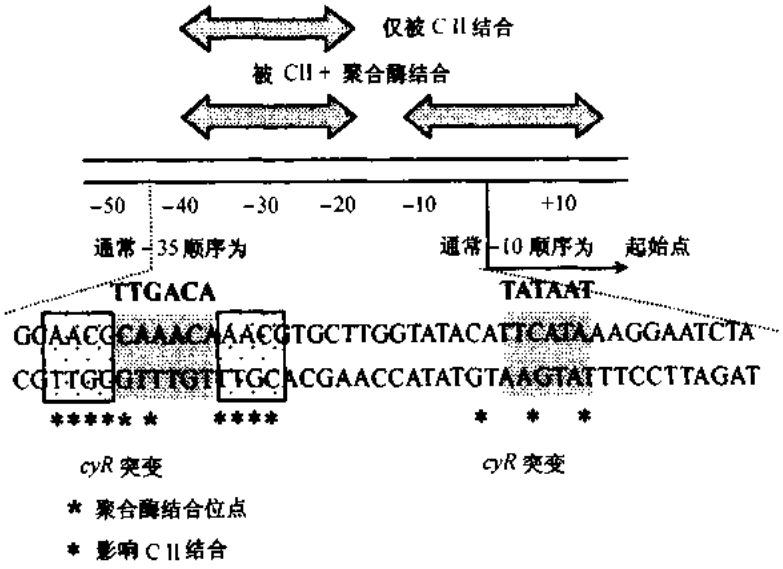


图 17-21 只有在 C Ⅱ存在时(结合在 -35 区域)RNA 聚合酶才和启动子 P_{RE} 结合。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 13.24)

-35 周围,分为两种类型,不仅影响到 RNA 聚合酶的结合,还会影响到 C Ⅱ的结合。在此区域的中心突变并不影响 C Ⅱ的结合,可能会阻断 RNA 聚合酶的结合。若突变发生在这个区域两侧的 4 聚体短重复序列上(TTGC)就会阻断 C Ⅱ的结合。此区域的 10 个碱基两端的 4 聚体是相同的。因此当 C Ⅱ识别两个四聚体,它位于 DNAα 的表面。

启动子的正控制意味着某些附属蛋白增加了 RNA 聚合酶起始转录的效率。表 17-1 表明无论是启动子和聚合酶之间相互作用的两个阶段还是其中的一个阶段(它们起始结合成闭合复合体(平衡系数以[K_B]表示)或是转变成开放复合体(以速率常数[K_2]表示))都能作为调节的靶子。

现在我们来看看在新感染噬菌体中溶原化是如何建立的。图 17-22 是早期阶段的概括,并表示 *c* Ⅲ和 *c* Ⅱ表达所产生的结果。*c* Ⅱ的存在有利于 P_{RE} 的转录,由这个转录本大量合成阻遏物,并立即结合于 O_L 和 O_R 上,阻遏物的结合直接抑制了 P_L 和 P_R 的一切转录,关闭了噬菌体所有基因的表达。C Ⅱ和 C Ⅲ的合成也被迫停止。它们是不稳定的 ,很快被降解。因此 P_{RE} 不能长时期地被使用。通过"建立"这个途径合成阻遏物也被终止了。但现在已有阻遏物结合在 O_R 上,使 CI 的合成由"建立"转换成"维持"途径,由启动子 P_{RM} 负责转录。虽然 P_{RM} 的功能是低水平的。但可以持续合成阻遏物。在打开维持循环的同时,关闭了高水平的"建立"循环,然后关闭了其他有关的功能。维持循环虽然仅低水平的转录 *cI*,但可足以满足溶原化的需要。

我们现在尚不知道建立溶原还需要其他什么功能,但 λ 感染时其 DNA 必须插入到宿主的基因组中。此插入需要基因 *int*(integrase)的产物。这个基因是由启动子 P_I 转录的,它表达时也需要 C Ⅱ蛋白。P_I 的序列与 P_{RE} 是同源的,它们都具有 C Ⅱ的结合位点。建立溶原控制循环所必须的功能是结合 DNA 的物理功能,这样溶原化的建立是在所有必须的事件同时发生的条件下进行的。

λ 的级联调控是十分复杂的,我们还不知道 C Ⅱ怎样以其他间接方式来启动溶原化。它从一个称为 P_{anti-Q} 的启动子发动转录,此启动子位于 *Q* 基因内。此转录本是 Q 区域的反义 RNA,可以和 Q 的 RNA 进行分子杂交,阻断 Q 蛋白的翻译。而 Q 蛋白又是裂解发育所必须的,*cI* 基因转录的结果不仅直接启动溶原化,而且也以同样的机制抑制 *cro*

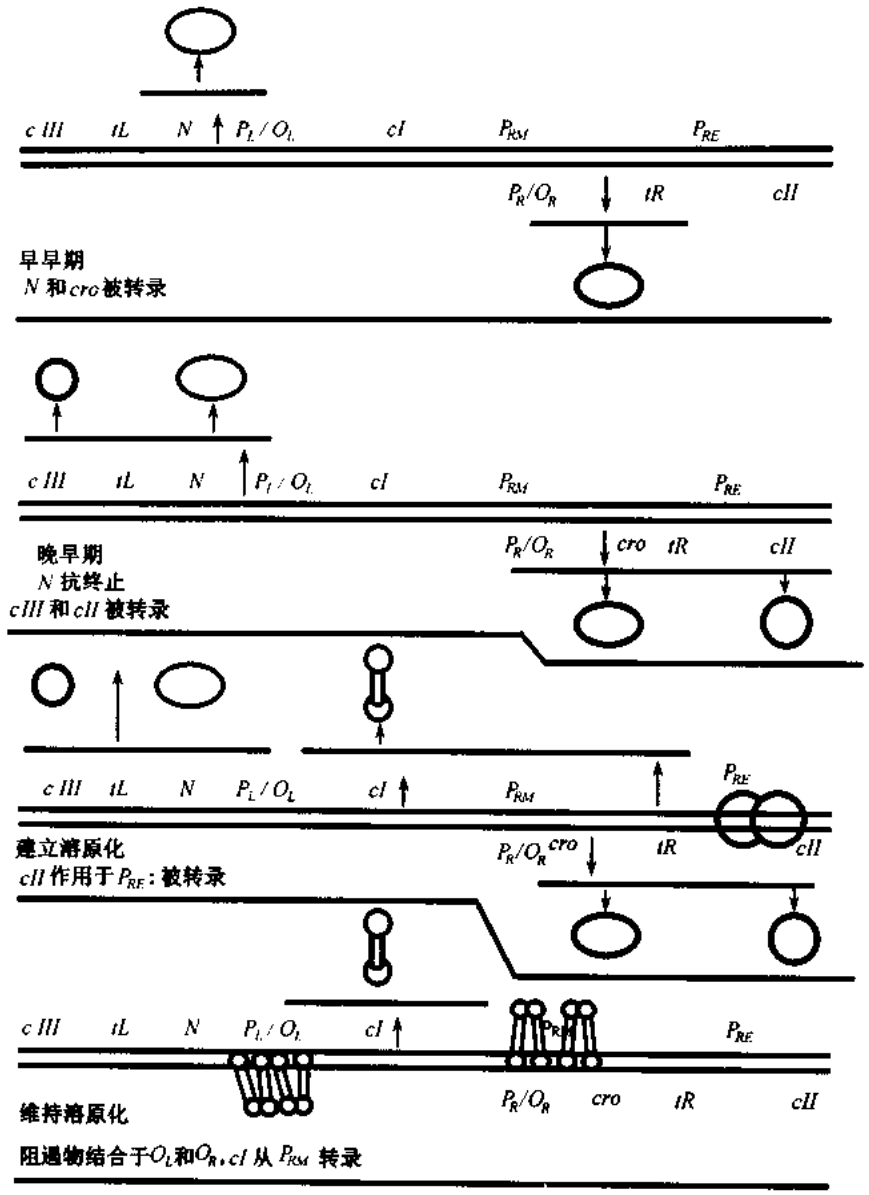


图 17－22　溶原化的建立和维持。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 13－25）

和 Q 的表达，从而间接地建立溶原化。这些调节基因对于抗裂解途径都是必要的。

第四节　Cro 蛋白的功能

本章的开始我们就谈到了 λ 的生命周期中有两种不同的途径：溶原化和裂解周期。溶原化是由建立一个自主调节的"维持"循环来起始的，对两点施加压力来抑制进入裂解周期。溶原化的建立实际上是在裂解级联调控的早期进行的。现在我们面临着另一个问题：噬菌体如何进入裂解周期的？

cro 基因编码另一种阻遏蛋白。它负责阻止 CI 蛋白的合成，其作用是关闭建立溶原化的可能性。*cro* 突变型通常是建立溶原化，而不进入裂解途径。这是由于失去了关闭 CI 蛋白表达的能力。

Cro 形成一种小分子二聚体（每个亚基为 9kDa），作用于免疫区，具有两种效能：①

阻断维持循环中阻遏物的合成，即阻止启动子 P_{RM} 转录；②也可抑制从 P_L 和 P_R 起始的早期基因的表达。这就意味着当噬菌体进入裂解途径时，Cro 负责抑制阻遏蛋白的合成及关闭早期基因的表达。Cro 结合于阻遏蛋白(CI)相同的操纵基因。Cro 含有一个 CI 相同的结构区域：螺旋 - 2 和识别螺旋 - 3 形成一个角度(这种结构的余下部分是不同的，表明螺旋 - 转角 - 螺旋基序是能在各种接触中起作用)。和 CI 相似 Cro 对称地结合于操纵基因上。

Cro 蛋白和 CI 蛋白在 G - T - H 区域的序列是相关的。此解释了它们能与相同的 DNA 序列结合的原因。Cro 与 DNA 的接触与 CI 相似，但只结合 DNA 的一个表面，它和 CI 不同，缺乏 N 端的臂，不能接触到 DNA 的另一面。

这两种蛋白作用相反，又怎样能作用于同一位点呢？答案是它们的亲和能力不同。它们分别结合于操纵元件上的不同位点(Cro 没有活性区域)。我们来看看 O_R，此是我们了解得比较多的位点，Cro 也就在此处执行它的作用。图 17 - 23 所示的一系列的事件(注意头两个阶段和溶原途径是相同的)。

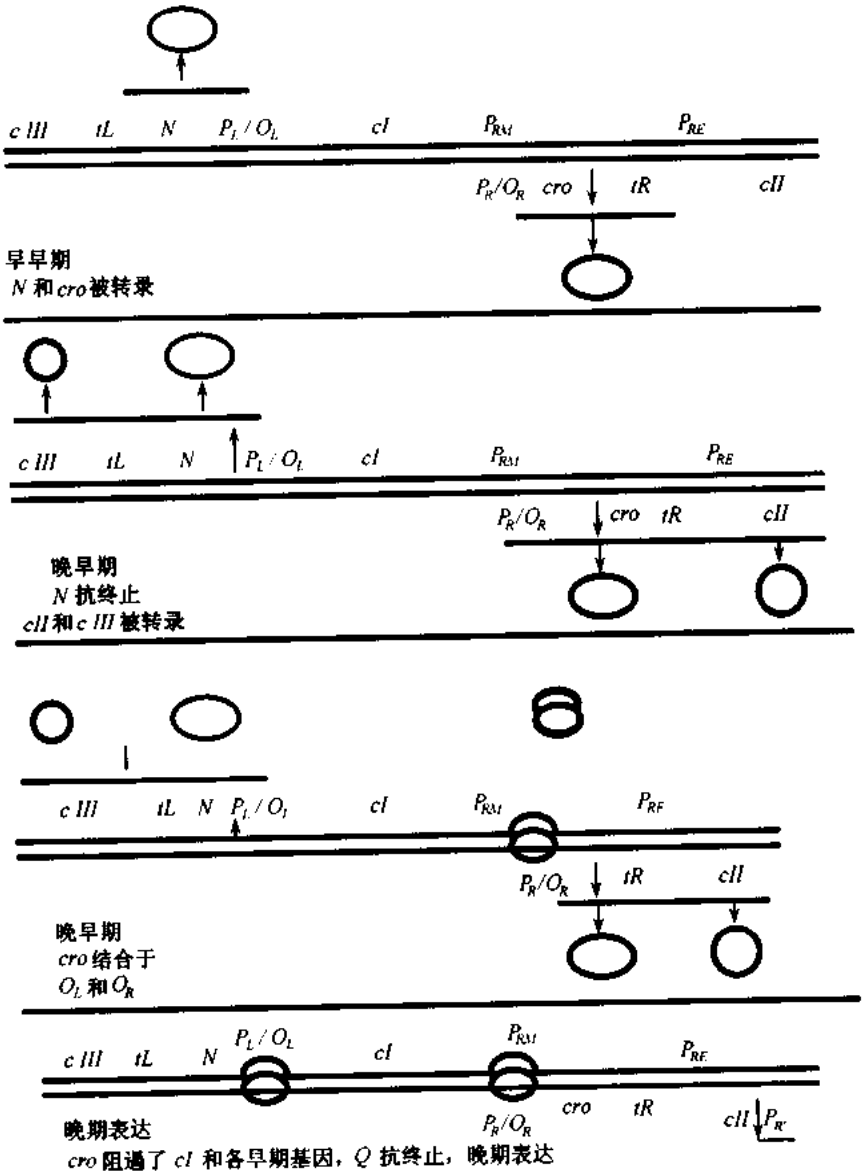


图 17 - 23　裂解周期中的早期表达和晚期表达的调节。(仿 B. Lewin: Genes V, 1994, Fig. 1 - 2)

Cro 对 O_{R3} 的亲和力要大于对 O_{R2} 或 O_{R1} 的亲和力，因此它首先结合于 O_{R3}，这就阻止了 RNA 聚体酶结合于 P_{RM}，这样 Cro 的第一个作用便是阻止溶原化的"维持"循环不能实现。然后 Cro 再结合于 O_{R2} 或 O_{R1}，这两个位点点亲和力是相似的，且没有协同效应。在这 2 个位点中 Cro 结合其中一个位点就足以阻止 RNA 聚合酶利用启动子 P_{R}。依次停止了早期功能的产物的产生。由于 C Ⅱ是不稳定的，P_{RE} 的使用也被停止。Cro 使 CI 蛋白的产生受到阻断。

就裂解周期而言，它是和 Cro 关闭早期基因的表达有关（虽然它并不完全消除早期基因的转录）。不完全作用是由于 Cro 对 O_{R1} 和 O_{R2} 的亲和力要此 CI 蛋白对 O_{R1} 和 O_{R2} 的亲和力低 8 倍。由于 pQ 的存在，直到早期基因的产物多少有些过剩时，Cro 才产生作用。但此时噬菌体又开始了晚期基因的表达，这些基因产物和噬菌体颗粒有关。

第五节　溶原化和裂解周期的平衡

溶原化周期和裂解周期关系十分密切，当噬菌体感染宿主时，二者之间的平衡决定了单个噬菌体的命运。CI 和 Cro 之间的对抗要么由建立自主的维持调控来打破僵局，要么关闭 CI 的合成进入晚期发育而告终（图 17－23）。这两种不同的结局如何决定呢？这两个途径的前 2 个阶段是相同的（图 17－22,23），它涉及到早早期基因的表达并进入晚早期基因的表达。所不同的是 CI 或 Cro 究竟哪一个来占据两个操纵原件。当 P_L 和 P_R 被抑制时，早期基因的表达将受阻，当 C Ⅱ和 C Ⅲ缺乏时，CI 通过 P_{RE} 的转录也将停止。关键的问题是 P_{RE} 是否在停止转录后接着 P_{RM} 激活，建立溶原化，还是 P_{RM} 不转录，而是 PQ 激活，其产物使噬菌体进入裂解发育。

建立溶原化的起始是 CI 结合 O_{L1} 和 O_{R1}。在此位点结合是很快的，然后通过协同效应阻遏物二聚体进一步结合于 O_{L2} 和 O_{R2}，这样就关闭 Cro 的合成，而开始了 P_{RM} 的转录。

进入裂解周期的起始是 Cro 结合 O_{R3} 上，阻断了从 P_{RM} 开始的维持溶原化的循环，然后 Cro 还必须结合于 O_{R1} 或 O_{R2} 及 O_{L1} 或 O_{L2}，阻断早期基因的表达。通过中断 C Ⅱ和 C Ⅲ的产生来导致中止 P_{RE} 合成阻遏物 C Ⅰ。当不稳定的 C Ⅱ和 C Ⅲ蛋白消失后，关闭了溶原化建立的发生。在溶原化和裂解周期之间发生转换的关键是 C Ⅱ。若 C Ⅱ是有活性的，通过 P_{RE} 启动子有效地合成阻遏物，C Ⅰ占据操纵基因。若 C Ⅱ无活性，阻遏物 CI 不能合成，Cro 就结合于操纵基因上。在任何特殊的条件下，C Ⅱ的浓度决定了感染的后果。C Ⅱ稳定性增加的突变使溶原化的频率也增加。这种突变发生在 C Ⅱ基因的本身或其他的基因上。C Ⅱ的不稳定性是由于它对一种宿主蛋白降解的易感性所产生的。它在细胞中的浓度受到 C Ⅲ的影响，也受到宿主功能的影响。

λ 噬菌体 C Ⅲ蛋白的作用是次要的：它可以保护 C Ⅱ免受降解。虽然 C Ⅲ的存在并不能保证 C Ⅱ的幸存，但在 C Ⅲ缺乏时 C Ⅱ总是失活的。

宿主基因产物的作用是通过下列途径。当宿主基因 *hflA* 和 *hflB*（*hfl*，high frequence lysogenization）突变时促进了溶原化。这种突变由于能使降解 C Ⅱ的蛋白失活而使 C Ⅱ稳定。宿主细胞对 C Ⅱ水平的影响为噬菌体两种途径的选择提供了一个因素。例如降解 C Ⅱ的宿主蛋白酶在营养丰富时被激活，那么噬菌体倾向于裂解细胞，若情况相反的话噬菌体倾向于溶原化。

此外两种途径的决定还和噬菌体的**感染复数**（multiplicity of infection，MOI）有关，即感染时噬菌体与细菌的比例。这是由于 C Ⅱ蛋白的单体不仅不稳定，而且是没有活性

的，只有以二聚体形式存在时才具有活性。当仅有 1－2 个噬菌体感染同一个细菌细胞时，产生的 C Ⅱ很少，其水平不足以形成二聚体，此时 λ 倾向于裂解周期；当感染复数提高时，C Ⅱ的浓度也随之增加，可以形成具有活性的 C Ⅱ二聚体，结果有利于向溶原化途径发育。

C Ⅱ蛋白不仅可以活化启动子 P_{RE}，还可以活化启动子 P_{I}。P_{I} 是负责 *int*(integrase)基因转录的。C Ⅱ蛋白的功能是双重的，可以促进阻遏物和整合酶的产生，有利于溶原化途径的选择。*int* 基因也可以由启动子 P_{L} 起始转录，可能是在缺乏 CI 的条件下才出现这种情况，但并不能产生整合酶。这是由于 P_{L} 转录的 int mRNA 会被细胞中的核酸酶所降解，而由 P_{I} 转录的 mRNA 就较为稳定，进一步翻译成蛋白质。这两种 mRNA 的差别就在其 3′端。由 P_{I} 起始的 mRNA 在 *int* 基因末端后面约 300nt 的终点子处停止转录。它具有典型的茎环结构，在其末尾有 6 个 U(图 17－24)。它对核酸酶是具有抗性的。而从 P_{L} 转录的 RNA，因为 RNA 聚合酶经过 N 抗终止蛋白的修饰可以通过终止子 t_{L2}，这样 RNA 得以延伸，形成了核酸酶的靶位点 *sib*。它首先被 RNase Ⅲ剪切，接着外切核酸酶从 3′－5′降解 mRNA，因此，这个位点是 *int* 基因下游的负调节位点，降解由下游向上游方向进行，故称其为**反向调节**(retroregulation)。当 *int* 编码序列的下游 *sib* 位点发生突变或者RNase Ⅲ基因发生突变(*rnc*⁻)，那么 P_{L} 转录的 mRNA 也不被降解，且可翻译成 Int 蛋白。

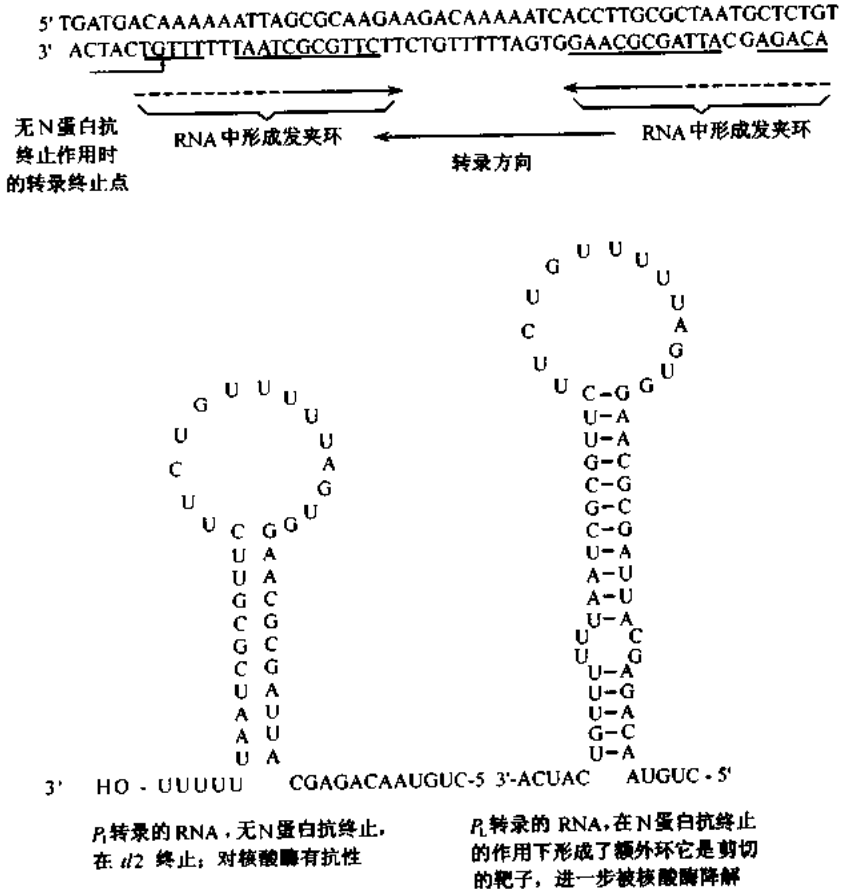


图 17－24 int 蛋白的反向调控的机制。(仿 J. D. Watson *et al.*: Molecular Biology of The Gene, 4th. ed., 1987. Fig. 17－20)

sib 突变阻止了 RNase Ⅲ 识别靶位点(图 17－25)，*rnc* 的突变具有相同的效果。P_1 转录的 mRNA 较为稳定正是其下游缺乏 *sib* 位点之故。

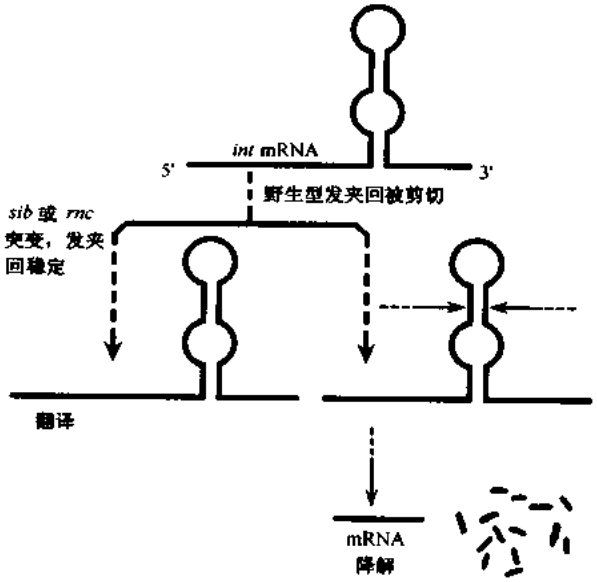


图 17－25　int RNA 的 3'端 sib 位点被 RNase Ⅲ剪切，导致 mRNA 的降解，结果产生反向调节。
(仿 B. Lewin: Genes V, 1994, Fig. 16－19)

第十八章

真核生物基因表达的调控

基本原理

在真核生物中基因表达的调节特点是①多层次，②无操纵子和衰减子，③个体发育复杂，④受环境影响较小

在真核生物中，RNA 聚合酶没有 σ 因子，不能直接识别启动子，必须依靠各种蛋白质因子帮助 RNA 聚合酶识别启动子，增加了调节的复杂性。

真核生物编码蛋白的基因含启动子元件和增强子元件，有的启动子元件是转录起始所需，有的起调节功能。他们可被特殊的调节蛋白识别和结合从而调节某些基因的表达。特殊的调节蛋白也和增强元件相结合，再通过和结合在启动子元件上的蛋白质相互作用来激活转录。增强子元件和启动子元件也与许多相同的蛋白结合。这表明两种调控元件是以相同的机制作用于转录的。

具有转录活性的染色质对 DNase Ⅰ的降解比非转录活性的染色质要敏感得多。这种敏感是转录激活区的 DNA 蛋白结构疏松的结果。对于有活性的基因而言其一定的位点称为超敏区，即对 DNase Ⅰ高度敏感区域。这些超敏位点可能是相应的 RNA 聚合酶和调节蛋白的结合位点。

大部分真核 DNA 在一定的碱基上进行甲基化，甲基一般都加在胞嘧啶上。在这些真核生物中，具有转录活性的基因比无转录活性基因甲基化的程度要低得多。

在真核生物中尽管没有发现操纵子，但在某些生化途径中相关功能的基因有时是连锁的。更多的情况是基因分数在基因组中，并列地受到调控。

在高等真核生物中，短期（short-term）基因调节系统是通过甾类激素来控制酶的合成。有些激素（如甾类激素）是以激素受体－复合体的形式直接和细胞的基因组结合来发挥它们的作用。另一些激素（如多肽类激素）作用在细胞表面上，激活表面受体系统，产生 cAMP 第二信使来控制相关基因的活性。激素作用的特异性是由于激素受体只存在于一定类型的细胞中，而且要通过受体复合物与细胞特异类型的调节蛋白的相互作用才能发挥功能。

真核生物基因表达能在 RNA 加工这一水平上进行调控。这种类型的调节包括从 RNA 前体分子到成熟的 mRNA 的过程，以至决定了其产物的类型。这种类型的调节可以通过选择不同 5′起始点，3′加尾位点以及不同内含子剪切产生不同的成熟 mRNA，最终合成不同的蛋白。

mRNA 从细胞核到细胞质的转运也是一个重要的控制点，剪接体滞留模型就反映了这一类调节的特点。在此模型中剪接体在前体 mRNA 上的装配与 mRNA 输出相竞争，这样在切除内含子时，在核中的剪接体用来固定 RNA，当所有的内含子被切除时剪接体也就就被解离，游离的 mRNA 能与核孔相互作用而输出。

基因的表达还可以通过翻译控制和 mRNA 的降解来进行调节。后者已成为基因 表达调节的主要控制点。mRNA 降解的速率和个体的结构特点有关。细胞因子和酶的作用与 mRNA 的稳定性也有一定的关系。

长期调节（long-term regulation）是在发育和分化过程中基因的激活和阻遏。

细胞周期的调控主要有两个关键的控制点 G_2 →M 和 G_1→S。很多的蛋白质因子都参与了调控，其中起主要作用的是 MPF（成熟促进因子），它由催化亚基 p34 蛋白和调节亚基周期蛋白（cyclin）组成。在细胞周期中催化亚基含量稳定，只有磷酸化和去磷酸化状态的变化。而 cyclin 是不稳定的，在 G_1 期合成 cyclin A，进入 S 期后降解，然后再合成 cyclin B 进入 M 期后降解。

p34 蛋白是一种蛋白磷酸化激酶，它在不同时期不同位点由其他的激酶进行磷酸化和去磷酸化，导致活性的改变。有活性的 p34 蛋白又可使一系列的蛋白磷酸化，并可激活某些转录因子，导致 S 期 DNA 的合成以及 M 期染色体的凝聚，形成了一个复杂的级联调控网络。

真核生物基因表达的调控要比原核复杂得多，特别是高等生物，不仅由多细胞构成，而且具有组织和器官的分化。细胞中由核膜将核和细胞质分隔开，转录和翻译并不偶联，而是分别在核和细胞质中进行。基因组不再是环状或线状近于裸露的 DNA，而是由多条染色体组成，染色体本身结构是以核小体为单位形成的多极结构。真核生物的个体还存在着复杂的个体发育和分化，因此真核生物的基因调控是多层次的，从 DNA 到染色质、到染色体。而原核生物基本上是在 DNA 这一层次上，在转录和翻译水平上进行调控。真核生物的染色体在正常情况下也存在丢失，扩增和重排，来进行"宏观"调控，在染色质的水平上受到组蛋白和非组蛋白的调节，有时要进行异染化或去异染化来调节活性；在 DNA 水平上和原核相似主要是转录和翻译两个水平的调控，但由于 5′端顺式元件的存在和多种转录因子的加盟，以及内含子的存在使调控更为扑朔迷离。

原核生物基因调控的中心是适应外部的环境条件，节省能量，维持生命为原则；调节的最主要形式是操纵子，真核生物由于长期进化的结果，受外界的影响比较小，但其本身的生理、生化的反应要求一系列复杂的调控，我们可将真核基因的调节分为两种不同的范涛：短期调节和长期调节（short－regulation and long－term regulation）。所谓的短期调节涉及到基因迅速地打开或关闭，以此对环境条件的改变或者细胞的及组织的生理条件的改变作出反应。如代谢的调节和一些应激反应。长期调节是生物的发育和分化所必须的，它在一定程度上受到程序化的控制，和外界环境的变化关系不大，而是受到内在的调节物质的控制。

基因的活性和基因的表达是两个不同的概念，基因的活性是指基因是否具备表达的条件，具有活性的基因是指它已具有表达的基本条件。但并不是说它正在表达；基因的表达是指这个基因已具备表达的充分条件，对其调节仅仅控制表达的水平。

第一节　染色体水平的调控

一、染色体丢失

有的生物在个体发育的早期在体细胞中要丢失部分染色体，而在生殖细胞中保持全部的基因组，被丢失的染色体上的遗传信息可能对体细胞来说没有什么意义，而对生殖细胞的发育是不可缺少的，现在我们来看下面的两个有名的例子。

马蛔虫（*Parascaris equoorum*）受精卵细胞内只有一对染色体（2n = 2）（另一亚种 2n = 4），但染色体上有多个着丝粒。在发育早期仅一个着丝粒发挥作用，保证正常有丝分裂的进行。发育后一阶段，在纵裂的细胞中染色体分成很多小片段，其中部分含着丝点，不含着丝点的片段在分裂中丢失。而在横裂的细胞中染色体并不丢失。第一次卵裂是横裂，产生上下两个子细胞，由于受精卵中含有的各种物质并不是匀一分布，而是从上到下呈一种梯度分布，也就是说上面的动物极和下面的植物极成分是不同的，它影响到分裂的方向和染色体的丢失。第二次卵裂时下面的子细胞仍进行横裂，保持着原有的基因组，而上面的子细胞却进行纵裂，丢失了部分染色体（图 18－1）。长此下去最下面的子细胞总是保持了全套的基因组，将发育成生殖细胞，其余丢失了部分染色体片段的细胞分化为体细胞。这可能是由于在植物极中有特殊的物质的存在，这种物质具有保护染色体不受削减的功能。

小表瘿蚊（*Mayetiole destructor*）受精卵的细胞核分裂，而受精卵不分裂，形成合胞

体(syncytium)。当核第3次分裂后,有两核移向卵后端的一个特殊区域,叫做**极细胞质**,在极细胞质中的核保持了全套的染色体(2n =40),将来分化为生殖细胞。而在一般的细胞质中,染色丢失了32条,只保留8条,将来分化为体细胞。若在细胞核移向极质前,用细尼龙丝在极质处进行结扎,不让核移入极质或者用紫外线照射极质,结果所有的核中都只保留了8条染色体。极细胞质中有很多核糖体样的颗粒,可以保护染色体不被削减,当紫外线照射后破坏了其功能,染色体照样也会丢失,最终不能发育为生殖细胞。

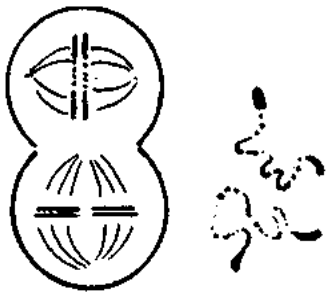


图18-1 马蛔虫染色体的丢失。

二、染色体的扩增

染色体的扩增其本质是细胞内特定基因拷贝数专一性大量增加可称为**基因的扩增**(amplification)。最为人们熟知的例子是两栖类和昆虫卵母细胞rRNA基因的扩增。非洲爪蟾的每条染色体上有约450拷贝编码18S rRNA和28S rRNA的DNA,它们成簇存在、重复串联在一起,形成**核仁组织区**。可是在卵母细胞中它们的拷贝数扩大了1000倍以上。卵母细胞十分特殊。它为胚胎的发育提供营养物质,直到发育成蝌蚪阶段。它的细胞质中也存在着翻译装置,是实行分化时必须的。

在卵子发生时,当内含物不断增加时,其体积也大大增加,形成了毗邻在一起的**滋养细胞**。(令人惊奇的是受精时卵中有很多核糖体,一个不带有18S和28S RNA的纯合体也可以发育到蝌蚪阶段,然后死亡。)卵母细胞核为细胞的第一次减数分裂的前期提供必要的物质,但进一步的发育却被制止了。两栖类减数分裂的染色体是巨大的,在活性区域中有很多的侧环,形成**灯刷染色体**。在卵母细胞的核中有数以千计大小不等的核仁。每个核仁含有大小不同的环状rDNA,它是染色体上18S和28S rDNA的串联重复单位经滚环复制从染色体上释放出来的。DNA的这种扩增的过程尚不完全清楚。此DNA的环产生rRNA,组装成核糖体,储备在核仁中到减数分裂时再释放出来。

另一种扩增的例子是在果蝇和其他双翅同昆虫的唾腺中。**多线染色体**(polytenne chromosomes)具有特异的膨松区,产生过量的DNA,而不是mRNA。George Rudkm早在1960年就指出:果蝇唾腺的多线染色体其常染色质DNA可以特异扩增上千倍,而在异染区附近只可复制几次。在染色体中存在着特殊的扩增位点,而其他的部分并不扩增。

在大部分体细胞中基因组的内容是保持不变的,但是某些顺序相对比例有时会发生变化。最有特色的例子是昆虫的幼虫发育,在它们的基因组中特定的序列会**过量复制**(coverreplication)或者**复制不足**(underreplication)。

复制不足发生在果蝇的多线染色体中。在此组织中,巨大染色体是由原来的二套染色体被多次复制而形成。黑腹果蝇(*D. melanogaster*)的基因组约有1/4的区域是由异染色区构成的,形成**染色体中心**(chromocenter),其大部分构成了卫星DNA序列。与二倍体细胞相比,在唾腺细胞中,异染色质的相对含量非常少。这是由于常染色质DNA顺序在多线化的过程中大量复制,而异染色体部分却不能大量复制。测定卫星DNA的含量表明此序列不复制或很少复制,而常染色质在唾腺中复制9次。

多线组织的rRNA基因也是复制不足的,rDNA的区域虽有6~7次的复制,但实

际上无论存在一个还是两个核仁形成区，最终达到的水平都是相同的。这种剂量控制形式可能仅限于 rDNA。

黑腹果蝇的基因组在多线组织中可以分成三种类型控制复制区。一个很自然的假设是各类区域的复制原点控制起始的数量。卫星 DNA 中的起始区不易识别；rDNA 中的起始区对 r DNA 基因的数量作出反馈性反应，停止其功能；一般的常染色质的起始区可持续起始直到结束。

在昆虫中特殊蛋白的编码序列存在着不同的扩增。在 *Rhyncosciara* 和其他昆虫的唾腺中产生 DNA 膨松(DNA puffs)。这种膨松表面上和双翅目中多线染色体存在的活性斑带相似，但膨松却含有局部的 DNA 扩增。此扩增的序列可能经过 4 次复制，使拷贝数增加了 16 倍。在黑腹果蝇的发育中，其卵壳基因提供了一种不同的扩增机制，卵壳蛋白是由多倍体的卵泡细胞合成和分泌的。昆虫的卵壳基因成簇排列，在黑腹果蝇中已鉴别出两组。在卵泡中其中一组位于 X 染色体上的基因在表达之前经 4 次重复，扩增了 16 倍；另一组基因在第 Ⅲ 染色体上，它们经 6 次重复，扩增了 64 倍。在这两组基因中，扩增区域延伸到卵壳基因另一侧约 45 ~ 50 kb。扩增最多的区域是在卵壳基因周围约 20kb 的区域。当扩增延伸到另一侧时，扩增显示出一个下降的斜率。

是什么在控制这种扩增呢？下降的斜率表明在单个分子中扩增区域的端点是异源的。从图 18 - 2 中的横型中可能得到解释。双向复制的重复起始发生在这个区域的中心部分。复制叉相距 10 ~ 50kb。这个模型认为整个的复制区作为一个**复制子**复制。此复制区是否含有一个顺式作用区负责复制呢？当这个区域的染色体发生突变时不会出现扩增。在杂合子中，染色体的突变也会阻止扩增。若这区域被易位到别处，它本身的复制叉也会发起扩增。

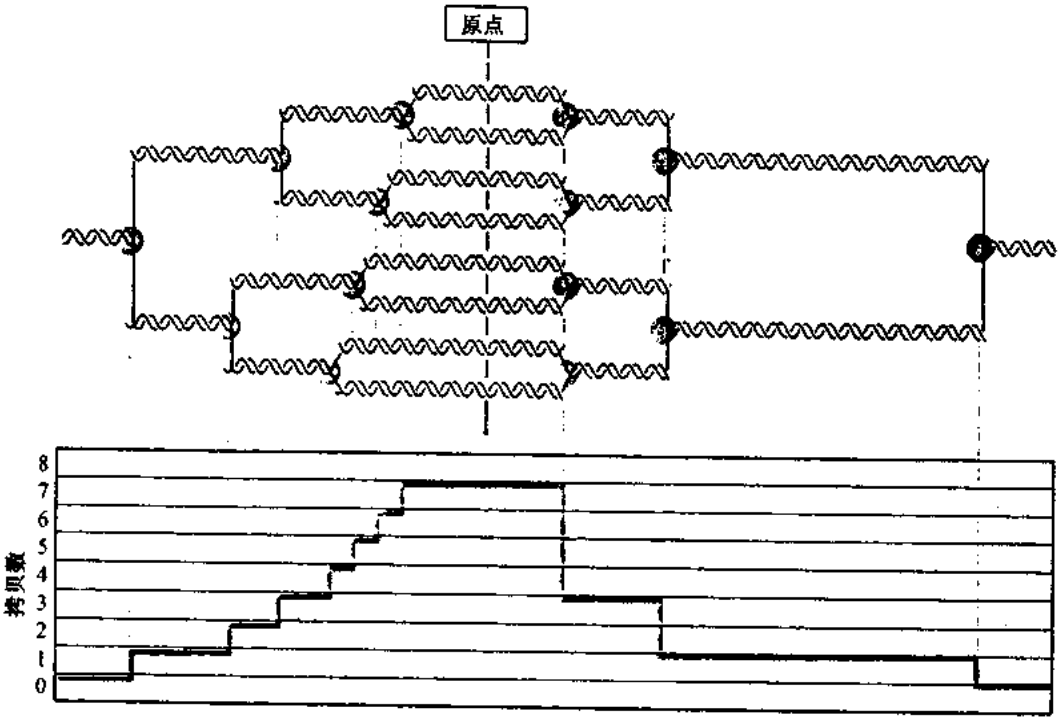


图 18 - 2　在单个复制子中局部区域通过多次起始来完成扩增。每个复制区都通过静态复制叉与侧翼的非复制区相连接。若复制叉在不精确的终止位点停止移动，则将会产生一个梯度，如图中左侧所示。若复制叉都在一个固定的终止位点停下，那么扩增的水平便突然减少，如图中右侧所示。(引自 B. Lewin: Genes Ⅳ, 1990, Fig. 35 - 1)

单眼缺乏(ocelliless)突变的性质和这种推测是符合的。这种突变导致了复杂的表型变化,它是约 3 个斑带区域的倒位,位于 X 染色体上,含有卵壳基因的区域。卵壳基因位于倒位左端 3kb 的区域。这个倒位引起重要的改变是扩增模式的变化。图 18－3 将野生型和小眼缺失突变的扩增区域进行了比较, 原来的扩增区在断裂点的左侧长 40kb;易位后扩增区在断裂点的右侧,长约 50kb。扩增的水平比原先有所减少,约是野生型极大值的一半(水平减少的原因尚不清楚)。

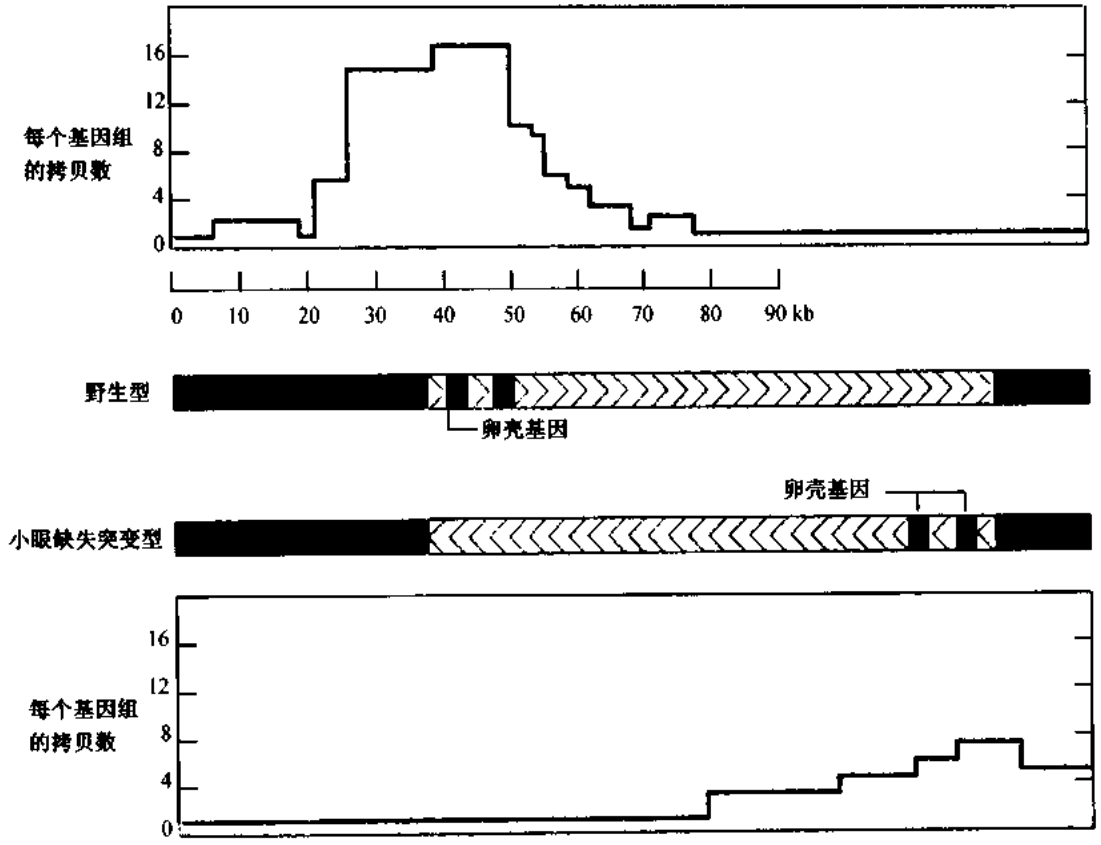


图 18－3 小眼缺失倒位改变了 DNA 的扩增区域。(引自 B. Lewin: Genes Ⅳ, 1990, Fig. 35－2)

结果表明复制原点能发起卵壳蛋白附近区域扩增。无论复制原点的位置发生什么变化,扩增区域的延伸仍保持不变。

三、基因组序列的选择性扩增

真核基因组具有调节内、外源基因组附加序列的能力。外源序列是通过转染(transfection)传导入细胞的。内源序列是由某一存在序列通过扩增而产生的。在这两种情况下它们的共同特点是:附加序列有可能以串联排列的形式含有很多重复单位的拷贝。包含在重复单位中的基因并不是每个拷贝都需要表达。但表达随着拷贝数趋向于增加。由转染或是扩增产生的多拷贝串联序列在细胞中可能以两种形式存在,即稳定系和不稳定系,(图 18－4)。若它形成染色体外的单位,则以一种不规则的方式遗传:在真核细胞分裂时,细菌的质粒是不能均等地分配到子细胞中,因此完整的单位很少存在。

内源序列的扩增是由那些对一定的试剂敏感的选择细胞所发动的。最好的例子是在一定的细胞系中加入氨甲喋呤（ methotrexate）结果导致扩增。此试剂是二氢叶酸还原酶（DHFR）的抑制剂，可阻断了叶酸盐的代谢。DHFR 的突变，改变活性使其对氨甲喋呤产生抗性。酶本身的改变是一种选择，酶的产量也可能增加。酶量的增加是 *dhfr* 结构基因数目的扩增所致。扩增发生频率要比自发的点突变的频率高得多，一般的范围是 $10^{-4} \sim 10^{-6}$。像这样类似的基因扩增现已发现 20 种以上。

在这些系统中大部分存在着共同的特点，就是简单地一次加入试剂并不能获得高度抗性的细胞，而是逐步地增加有毒试剂的剂量，使细胞渐渐适应，从而产生抗性细胞。因此基因的扩增是需要一系列的步骤。扩增也仅仅只发生在二个 *dhfr* 等位基因中的一个上。对氨甲喋呤抗性的增加是依靠这一基因座位扩增的程度来完成的。在抗氨甲喋呤细胞中 *dhfr* 基因数目的变化范围是 40～400 拷贝，拷贝数的多少取决于选择的程度和单个细胞系本身。*mtx*（metnotrexate 抗性）细胞系分为稳定系和不稳定系两类，细胞在缺乏氨甲喋呤的条件下生长时除去原来的高水平 DHFR 活性的压力，它们的反应不同，二者的差别如图 18－4 所示。

（1）在稳定系中搁增的基因得以保持，因为它们位于染色体上 *dhfr* 基因的位点上，而其内源染色体上的等位基仍保持正常的单个拷贝。

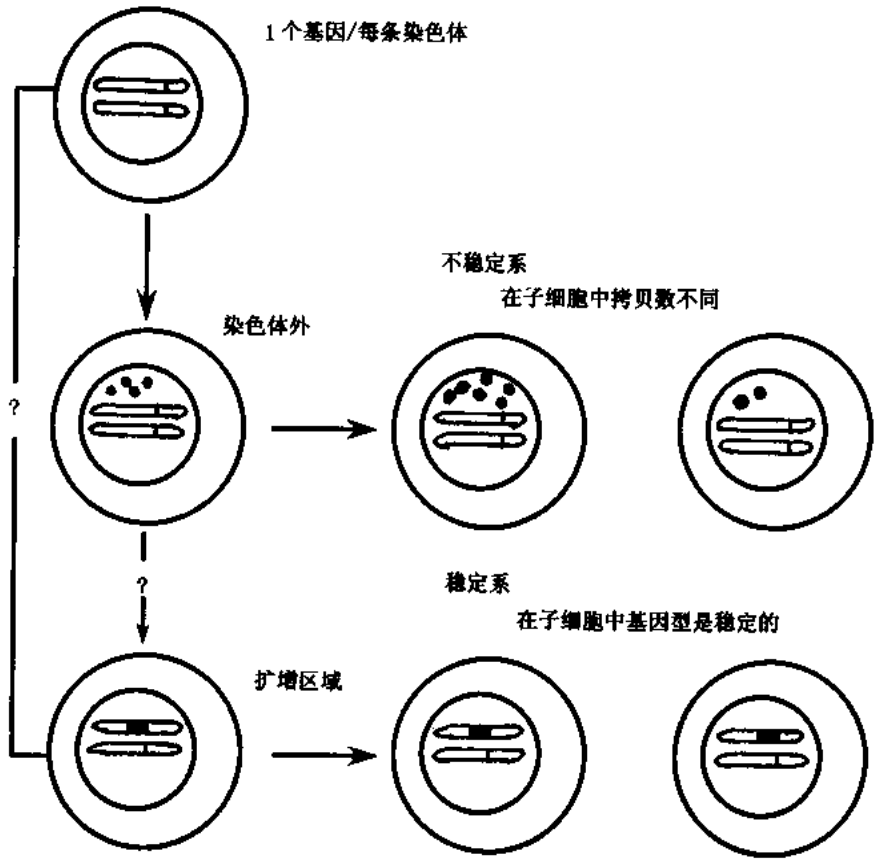


图 18－4　*dhfr* 基因扩增的两种类型。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32－27）

（2）在不稳定系统中，当选择压力释放时，扩增的基因会部分丢失。未丢失的基因是以染色体外的排列方式存在。

基因扩增在染色体上有一个可见的效应。在稳定系统中，*dhfr* 座位能形成一个均**染色区**(homogeneously staining region，HSR)，例如图 18－5 所示的情况。称其为均染区是由于在染色体上存在着一条增加的区域。用分带技术来处理(如 G 带)后，这一区域不显示任何染色体带，表明带间的某一区域经历了一种扩增。

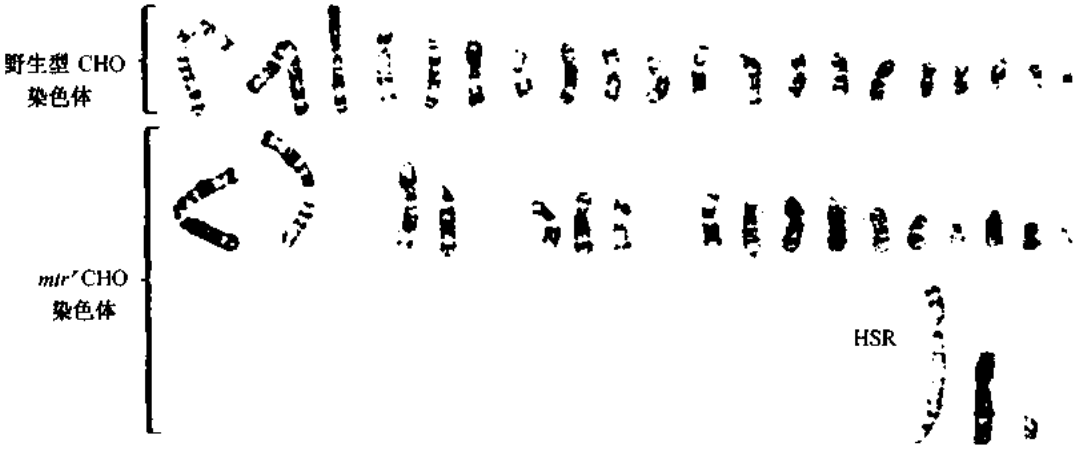


图 18－5　二－氢叶酸还原酶基因（*dhfr*）的扩增拷贝在染色体上产生了一个匀染区(HSR)。(转引自 B.Lewin:Genes Ⅵ，1997，Fig.32－28)

在不稳定的细胞系中，携有 *dhfr* 基因的染色体没有什么明显的改变，但是可以看到很多染色体外的点状染色体，称为双微体(double-minute chromosomes)(图 18－6)。在典型的细胞系中，每个双微体带有 2～4 *dhfr* 基因。双微体可以复制；但它们没有着丝粒，结果它们不能附着在有丝分裂的纺锤丝上，因此不能均等分离进入子细胞。虽然它的名字叫双微体，实际上可看作为染色体外的**微小染色体**。

双微体的不规则遗传解释了在这个细胞系中氨甲喋呤抗性的不稳定性 。在细胞分裂时双微体连续地丢失；在氨甲喋呤存在时，*dhfr* 基因减少的细胞将会死亡。只有哪些保持足够数量双微体的细胞才能幸存下来。双微体的存在减少了增殖细胞的比例。

图 18－6　被扩增的染色体外的 *dhfr* 基因产生了双微体染色体，如照片上所显示的白色小点。(转引自 B.Lewin:Genes Ⅵ，1997，Fig.32－29)

当去除选择压力时，缺乏基因扩增的细胞便有了一种优势，它们可以迅速产生后代，在群体中占主导地位。这就解释了为什么这种扩增状态仅存在于有氨甲喋呤的环境中。

由于双微体不均等分离，在分裂时受到选择的细胞中，其拷贝数能迅速增加。在培养条件中增加氨甲喋呤的水平，结果带有大量双微体的细胞也增加。双微体的行为解释了 *mtx'* 条件的逐步进化和不稳定细胞系中 *dhfr* 基因水平持续不断地波动。

稳定和不稳定细胞系，两者都是在氨甲喋呤抗性长期选择之后产生的。基因扩增的最初阶段是由什么起始的呢？在短期选择之后，大部分或全部的抗性细胞都是不稳定的。很明显染色体外的拷贝的形成要多于在染色体内的扩增。在此过程的最早阶段，扩增的 *dhfr* 基因可能是作为小的染色体外的单位存在，这在双微体或其他染色体变化之前是可以检测出的。扩增的区域要比 *dhfr* 本身长一些。这个基因长约 31kb 但在染色体的 HSR 中重复单位的平均长度是 500～1000kb。扩增区域的延伸在各细胞系是不同的。在双微体中含有 DNA 的总量的范围是 100～1000kb。

染色体外的拷贝是怎样产生的呢？我们知道它们后代产生时并未丢失原来染色体的拷贝。这有两种可能：①复制的附加环可能在 *dhfr* 基因附近起始，然后通过 dhfr 拷贝之间的非同源重组而产生（图 18－7）；②或是等位基因之间的非同源重组来起始这个过程。额外的拷贝可能是通过类似于重组的事件将其从染色体释放出来。根据这些事件的特点，不难推测，它可能会产生出一个含有单个或多个拷贝的 DNA 分子。若双微体含有环状 DNA 的话，它们之间的重组似乎可能产生多体分子。

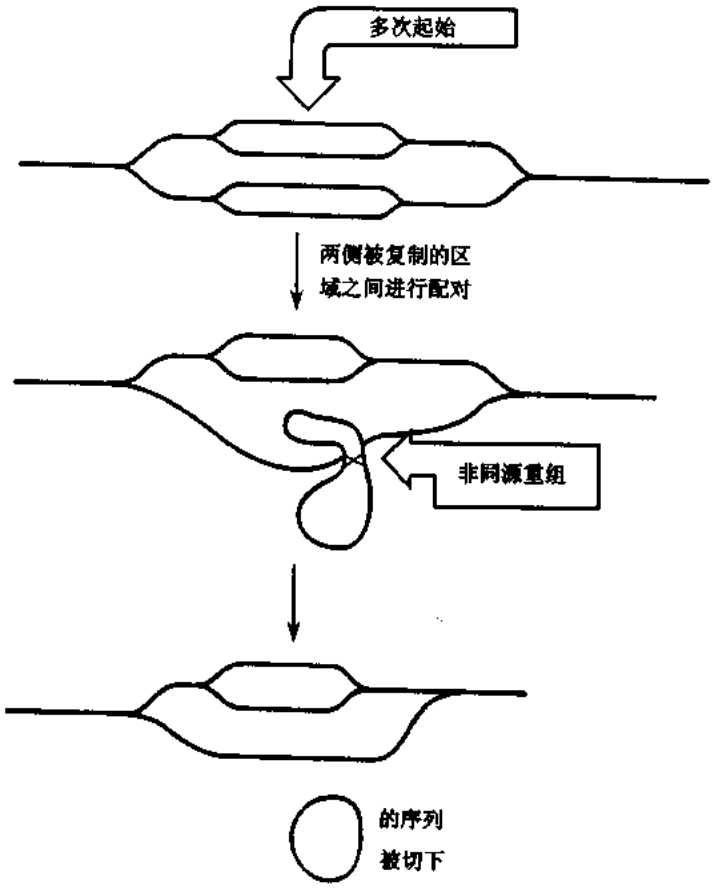


图 18－7　染色体外的 DNA 拷贝可能是通过复制子之间的非同源重组产生的，而这些复制子又可能经多次起始复制所形成的。（引自 B. Lewin: Genes Ⅳ, 1990, Fig. 35－6）

一旦染色体外的基因被扩增了，细胞状态的改变就由这些基因介导，而不由原来染色体上的拷贝来控制。当除去后，此细胞系以通常的方式丢失了双微体。若再加氨基喋呤，正常的基因又被扩增，产生一个新的双微体群体。这表明没有一个突变基因的染色体外的拷贝整合入染色体。

这些结果还表明突变系的双微体只带有突变基因。若每个双微体多于一个 *dhfr* 基因时，那么所存的双微体必定表现为突变型。这表明多拷贝双微体可能由单个的染色体外基因产生的。

细胞系产生 HSR 扩增的另外一些情况是对氨甲酰基转移酶(transcarbamylase)抑制物的抗性，它是通过 CAD 基因的扩增来实现的(CAD 是一种蛋白，它是 UMP 合成途经中三个酶中的第一个)。CAD DNA 的扩增是在染色体中发现的。在此情况下，被扩增的基因是以分散的形式分布在扩增区域，常涉及到一条以上的染色体。

四、染色体的重排

在原核生物中沙门氏细菌的相转变和 Mu 噬菌体 G 片断的倒位都属于基因组的重排。而在真核生物中染色体的重排更为复杂，涉及到很多蛋白质因子的级联调控。最典型的例子是免疫球蛋白在成熟过程中的重排以及酵母的交配型转变，由于内容较多又较为重要，下面另立一节来进行系统地介绍。

第二节 真核生物中染色体的重排

一、酵母的变配型的转变

1. 酵母的变配型

在很多真菌中其有性生殖的过程都需要不同**交配型**(matimg-type)的菌株相互接合才能产生二倍体的合子。相同的交配型之间是不能接合的，如酵 母的 a 型和 α 型(图 18－8)。细胞的交配型是由 *MAT*(matimg) 座位的遗传信息决定的。在此座位上带有 *MAT* 等位基因的细胞就叫做 a 细胞；同样带有 *MAT*α 等位基因的细胞就称为 α 细胞。只有 *a* 型和 α 型细胞之间才能交配，相同型细胞之间是不能交配的。

不同交配型细胞之间的识别是由分泌的**外激素**(pheromones)决定的。α 细胞分泌小分子的多肽 α 因子 (表 18－1)；a 细胞分泌 a 因子。α 因子是一个 13 肽(WHWLQLKPGQPMY)；a 因子是 12 肽(YIILGLFWAPY)，通过加法呢酯(farnesyl)(酯类)而被修饰及羧基甲基化。这两种肽的前体都经过剪切，最后释成成熟的肽序列。

一种交配型的细胞带有另一种交配型外激素的表面受体。当两种不同交配型细胞相遇时，它们的外激素与对方的受体相互作用，使细胞周期停止在 G_1 期，并发生各种形态学变化，包括产生胶着。在有效的接合中，细胞周期的停止使细胞和核得以融合，产生一个 α/a 二倍体细胞(图 18－8)。

α/ a 细胞带有 *MAT*α 和 *MAT*α 等位基因，并具有和单倍体细胞不同的特点 (表 18－1)，尤其是 α/a 细胞能形成孢子，而单倍体细胞不能形成孢子。

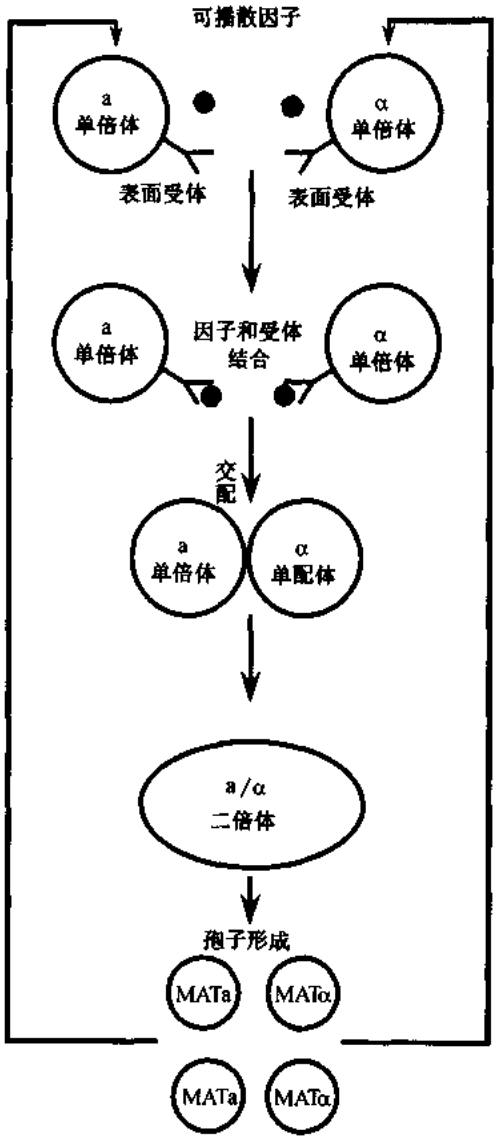


图 18－8　酵母的生活史：单倍体 MATα 和 MATα 交配，产生杂合的二倍体。
（仿 B. Lewin: Genes V, 1997, Fig. 32－1）

通过突变可使 a 或 α 细胞失去接合的能力，据此我们可以推断出一些酵母交配型途经的信息。通过这种方法鉴别出的基因称 *STE*（不育 sterile），*STE2* 基因是在 a 型细胞中，而 *STE3* 是在 α 细胞中的，其他 *STE*（1、4～7、11、12、18）基因是两者共有的。

接合是一种交互的过程，一种类型的细胞分泌的外激素与另一种类型细胞所带有

表 18－1　交配型控制的各种活性

	MATa	MATα	MATa/MATα
细胞型	a	α	a/α
交配	是	是	不
孢子形成	不	不	可
外激素	a 因子	α 因子	无
表面受体	结合 α 因子	结合 a 因子	无

B. Lewin: Genes V. 1994, Table 36.1

的受体相互作用。外激素和受体由不同的基因编码。

2. 交配型的信号传递

两种因子中任何一种因子与受体相互作用都可以接通反应途经，因此共同途径中某一步骤的突变不可能对两种细胞都起相同的作用。如 a 因子发生突变，只影响到 a 和 α 细胞受体的相互作用而不能接合，但不影响 α 因子和 a 细胞受体的相互作用。通过这种途径仍可以进行接合。

图 18－9 表示交配型反应中的下一个步骤涉及到细胞膜上的蛋白，它们和 G 蛋白（一种普遍性的膜受体蛋白）有关，G 蛋白含有 α，β 和 γ3 个亚基，α 亚基结合鸟苷，当此鸟苷水解为 GDP 时，α 亚基就和 β、γ 形成复合体；当 GTP 取代了 GDP 时，G 蛋白被激活，α 亚基从 β－γ 二聚体中释放出来。亚基的分离使 G 蛋白激活了信号传递途径中的下一个蛋白。由于这种相互作用，信号将有效地从 G 蛋白传递到靶蛋白，这一过程就称为信号传递（signal transduction）。细胞表面接收了信号后，能通过这种途经将信号传递到细胞内部。

信号传递的普遍类型是 α 亚基的激活，再与靶蛋白相互作用。但在交配型途径中情况不同。由 βγ 二聚体激活了下一个步骤。交配型途径中的 G－三聚体的成分通过基因突变已鉴别出来了，那就是 *SCG1*，*STE4* 和 *STE18*，它们的作用是负责和外激素结合。*SCG1* 编码 Gα 蛋白，它失活后导致了外激素的组成型表达，这是由于 Gα 在失活三聚体中不能维持 Gβγ。这种突变是致死的，因为其作用导致细胞周期的停止。*STE4* 基因编码 Gβ 蛋白，它的失活或者编码 Gγ 蛋白的 *STE18* 基因的失活不能激活下一步的蛋白，这样则不能进行起交配型反应，从而导致不育。

在后面的一些步骤也要涉及到其他的基因（图 18－10），*STE11* 和 *STE7* 的产物涉及到激酶的级联调节，它们的磷酸化能激活下一个蛋白。STE12 是一个 DNA 结合蛋白，它是一种转录因子，其产物对接合是必要的。

图 18－10 中没有把各种 STE 基因都放到极联调节中，这些成分的顺序也不是一定的。这些基因如编码激酶的 Fus3 和 Kss1 可能属于级联调节内或存在于分枝上。原理是清楚的：由外激素和受体相互作用产生的信号，沿着级联调节传递，激活接合所需功能。

（1）交配型的转换

某些品系的酵母具有转换（switch）交配型的能力，即从 a 型变成为 α 型，或从 α 型转变为 a 型，它们是如何转换的呢？这些品系带有显性等位基因 *HO*（表 18－2）并频繁地改变它们的交配型（常常每代改变一次），带有隐性等位基因 *ho* 的品系有一个稳定的交配型，其交配型改变的频率仅有约 10^{-6}。

HO 的存在导致酵母群体基因型的改变。转换的存在表明所有的细胞都含有 *MATa* 和 *MATα* 型的潜在信息，但只表达一种。交配型的信息在何处呢？转换成 *MATα* 型是需要 *MATa* 型，同样转换成 *MATa* 型也需要 *MATα* 型。它们在同一条染色体上，*HML* 位于 *MAT* 的左边，*HMR* 位于 *MAT* 的右边。

1977 年 Hicks，T.B. 等对酵母交配型的转换提出了暗箱模型（cassette model）（图 18－11），提出 MAT 是有活性的暗盒（active cassette），可以是 α 型，也可以是 a 型。HML 和 HMR 是 MAT 左右两侧的同源区，是沉默暗箱（silent cassettes）。通常 HML 带有 α 暗盒，而 HMR 带有 a 暗盒，所有的暗盒都带有编码交配型的信息，但只有 MAT 是活性的

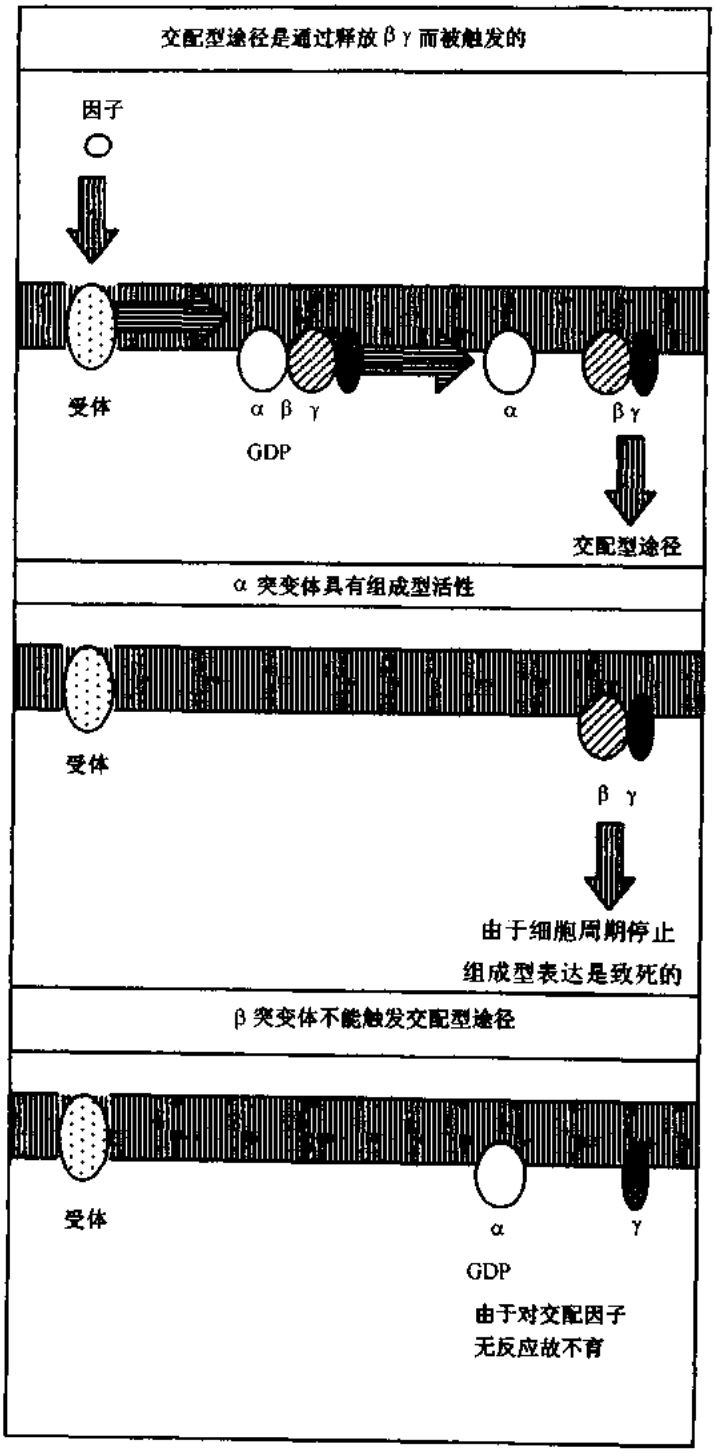


图 18-9　a 或 α 因子与受体相互作用触发激活了 G 蛋白，其 βγ 亚基将信号转递给下一个蛋白。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32-2）

暗盒，可以表达，*HML* 和 *HMR* 都不能表达，故称为沉默暗盒。当活性暗箱被沉默暗箱信息所取代时就发生了交配型转换。新"装进"活性暗盒的信息就可以表达。

交配型转换是非交互的，在 *HML* 或 *MMR* 上的拷贝替换了 *MAT* 上的等位基因。若 *MAT* 发生突变，当转换时它被取代掉而且永久消失，新换上的拷贝并未发生改变，表

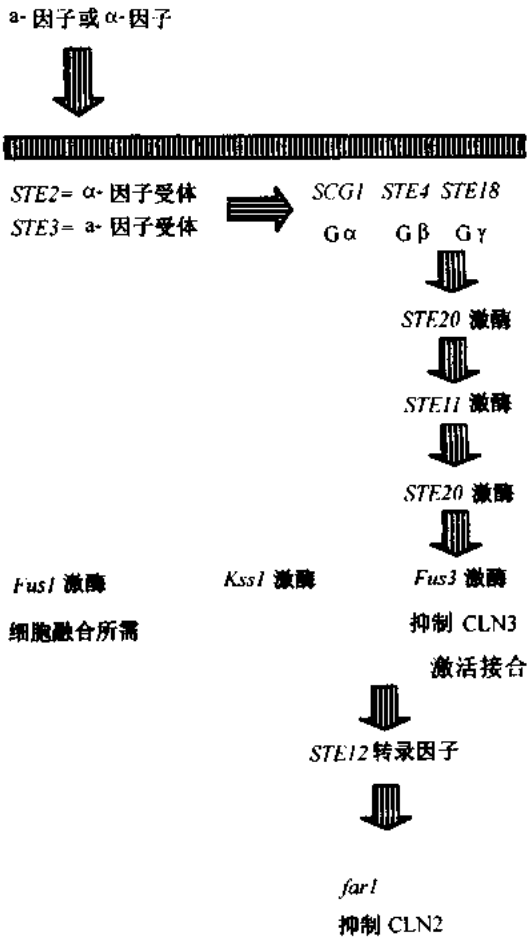


图 18-10 相同的交配型反应是通过外激素与其受体相互作用而引起的。此信号是通过一系列激酶转递给转录因子，在再执行最终的功能。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32-3)

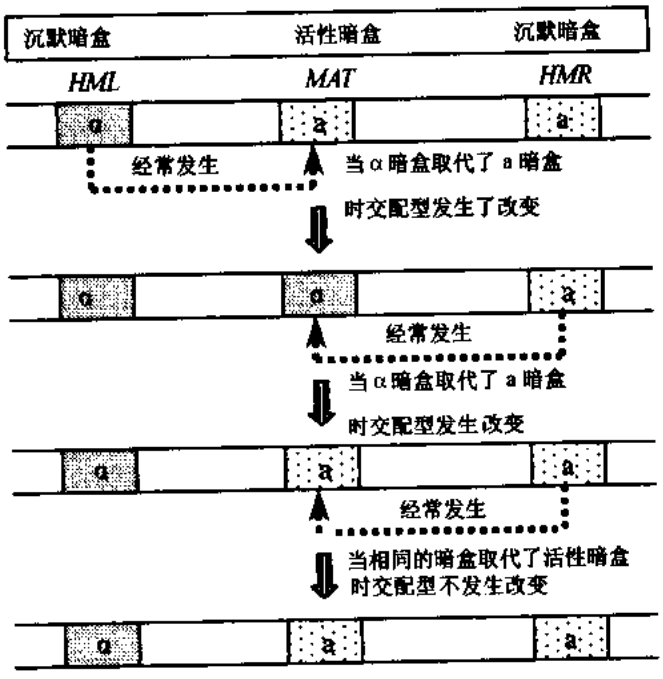


图 18-11 酵母交配型转换的暗盒模型。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32-4)

明了非交互转换。

在 *HML* 和 *HMR* 上的拷贝也会发生突变，在这种情况下转换可将突变的等位基因导入 *MAT* 位点。但在 *HML* 或 *HMR* 仍保留着一份突变的拷贝，这很像转座中的复制（见第 25 章），供体元件在受体位点产生了一个新的拷贝，而其本身仍保留在原来的位置上。

交配型转变是一个直接的事件。在转变中仅有一个受体（*MAT*），但有两个供体（*HML* 和 *HMR*），转换通常涉及到 *MAT* 上的拷贝被 *HML* 或 *HMR* 上的拷贝所取代。80% ~ 90%的转换，是 *MAT* 位点上的等位基因被另一个相对交配型等位基因所取代。通过细胞表型的决定显示了这些基因的作用。a 型细胞优先选择 *HML* 为供体；α 型细胞则优先选择 *HMR*。有 5%的细胞被同类型新拷贝基因所取代；还有约 5%并不发生变化。

(2)沉默暗盒和活性暗盒的结构

好几组基因涉及到交配型的建立和转换（表 18 - 2）。有的基因实际上涉及到交配型，它们包括抑制暗盒表达的基因，转换交配型或在接合中执行有关功能的基因。

比较一下沉默暗盒（*HMLα* 和 *HMR*a）和活性暗盒（（*MAT*a 和 *MATα*）的序列，我们就能描绘出决定交配的序列。图 18 - 12 所示每种暗盒都有共同的序列，侧翼序列 *W*. *X*

表 18 - 2　酵母接合作用有关的基因及功能

基因		产物及功能
asg	*MFa1 MFa2*	a-因子
	STE2	α-因子受体
	STE6	分泌 a-因子所需
	SsT1	分解 α-因子（蛋白分解酶）
asg	*MFa1 MFa2*	α-因子
	STE3	a-因子受体
	STE13	α-因子加工
hsg	*SCG1*	G 蛋白 α-因子
	STE1	G 蛋白 β-因子
	STE18	G 蛋白 γ-因子
	STE5	活化 *asg*·*asg*
	STE12	转录因子
	STE7、*STE11*	激酶（蛋白磷酸化酶）
	Fus3	激酶，关闭细胞周期，开通接合
	Kss1	激酶
	fus1	细胞融合
	far1	关闭细胞周期
	Kar1	核融合
	SIR	阻遏沉默暗盒的表达
	HO	内切酶，起始交配型转换
	SIN1 - 6	阻遏 *HO* 表达
	SWI1 - 5	为 *HO* 表达导致转换所需
	RME1	减数分裂和孢子形成的抑制因子
	HML/HMR	储存沉默的交配型信息

asg：a 交配型特异基因群，*asg*：α 交配型特异基因群，*hsg*：单倍体特异基因群。

和 *Z* 夹着一个中心区 *Y*。在 a 和 α 型暗盒中 *Y* 区域是不同的，分别称之为 *Ya* 和 *Yα*。中心区的两侧区域本质上都是相同的，仅 *HMR*α 缺少了 *W* 区。沉默暗盒是不表达的，活性暗盒可以由 Y 区中的启动子处转录。

***SIR* 基因关闭 *HMLα* 和 *HMRa* 基因** 原来认为 *HMLα* 和 *MHRa* 之所以"沉默"是由于它的缺乏表达所必须的启动子位点，其实不然，因为 *MATα* 和 *MAT*a 特异性 mRNA 的转录是在 Y 片段的内部起始的，*MAT* 座位和沉默基因必定具有激活它们所必需的相同启动子。差不多在同一时期遗传学家们发现了 4 个不连锁的基因沉默交配型调节基因，*SIR*(silent information regulator) *1*、*2*、*3* 和 *4*，这些基因产物共同起**反式作用**(它们并不在同一条的染色体上)来阻止沉默交配型基因的表达。若这 4 个 *SIR* 基因中任何一个基因失去作用的话，那么 *HMLα* 和 *HMR*a 基因就会以 *MAT* 座位上的基因相同的速率进行转录。为了寻找 SIR 产物的作用位点，首先通过体外诱变在 *HMR*a 和 *HMR*α 中产生缺失，然后利用 DNA 转化的方法将缺失的 DNAs 插入合适的酵母菌株中。

***E*$_L$(near *HML*) 和 *E*$_R$(near *HMR*)** 缺失突变分析表明在各个 *HML* 和 *HMR* 的

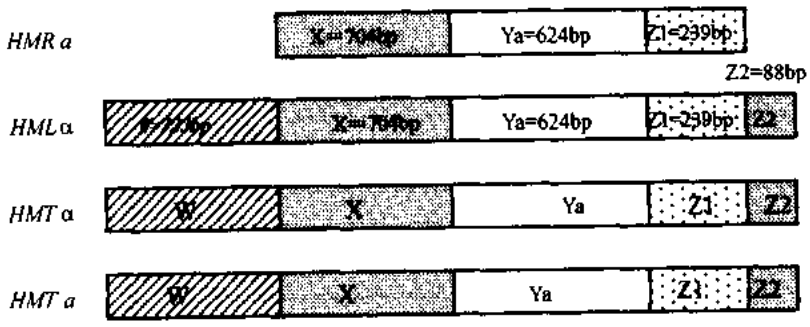


图 18－12 沉默暗盒和活性暗盒结构相同，但 HMRa 缺乏侧翼序列。
(参考 J.D.Watson *et al*:Molecular Biology of the Gene,4th ed.,1987,Fig.37－4)

上游具有阻遏它们表达的位点，此靶位点有时称为 **E 沉默子**和 **I 沉默子**，或 *E*$_L$(near HML)和 *E*$_R$(near *HMR*)。这些控制位点具有 2 个有趣的特点：① 它们具有负的增强子的作用，它们能对远隔 2～5kb 的启动子发挥作用，无论在靶启动子的上下游两个方向都可起作用，所以它们被称为**沉默子** (silencers)；② 它们和可能具有复制起点作用的 ASR 序列相联(图 18－13)。

温度敏感 *sir* 突变体的表达被用于研究 *HML* 和 *HMR* 的阻遏与 DNA 复制之间的关系。带有 *sir*^{ts} 突变的细胞在高温下培养，HML 和 HMR 得到表达，然后降低温度使 SIR 活性恢复，当细胞 G_1 期被阻断时，HML 和 HMR 仍持续表达。但当 DNA 复制时，HML 和 HMR 就被抑制了，这表明与 E 元件相连的 ASR 序列对于建立 *HML* 和 *HMR* 的阻遏是必须的。这是一种阻遏的新形式，现在我们还不知其分子机制是怎样的。

暗盒的活化导致了染色质结构的改变，所有的暗盒都具有 DNase I 的超敏感位点，但在活化和沉默暗盒之间存在着重要的差异。核小体结构及超螺旋的程度都有改变，可能和基因的激活有关。

图 18－14 表示活化暗盒和沉默暗盒的超敏感位点。*MAT*a 和 *MAT*α 中的超敏感模式一般是相同的。超敏感位点是在启动子及表达区域两侧中的一侧。*HML* 和 *HMR* 几乎没有这些位点但诱导产生的 *sir*$^-$ 突变恢复了所有丢失的超敏感位点。因此暗盒的

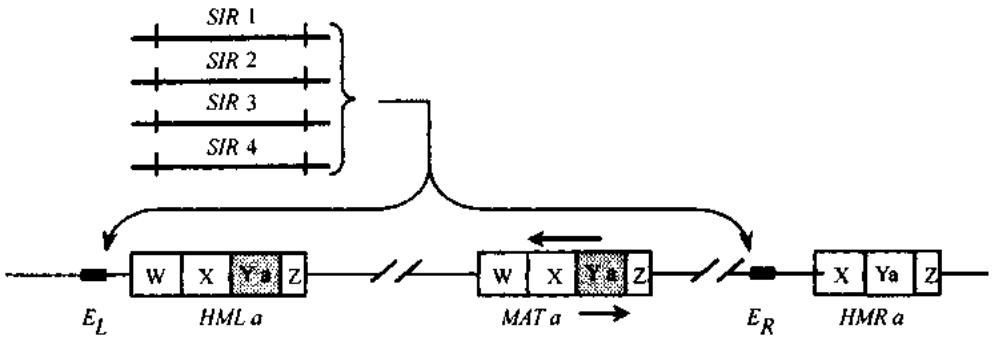


图 18－13　SIR 蛋白作用 E 区，关闭 HMLα 和 HMRa 的表达。
（参考 J. D. Watson *et al.*：Molecular Biology of the Gene，4th ed.，1987，Fig. 37－4）

活化和超敏位点的存在有联系。

4 个 *SIR* 座位之间的关系是什么呢？我们还不知道它们是编码单个调节蛋白的亚基，还是编码不同蛋白相互作用发挥功能或是单独地和靶位点起作用。虽然看来其中还没有一种蛋白结合 DNA，但它们的存在改变了染色体的结构。

由于组蛋白 H4 的 N－端的缺失可激活沉默暗盒，故暗盒是否有活性取决于其染色质结构。在 H4 的 N－端少数几个氨基酸的突变具有相同的激活作用。这种突变的效应可能被 *SIR3* 中产生的新的突变所掩盖，也可能被 *SIR1* 的过度表达所掩盖，表明 H4 和 SIR 蛋白之间存在特殊的相互作用。这种相互作用的特点尚不清楚，但它可能和 H4 的乙酰化及其他化学修饰的影响有关。

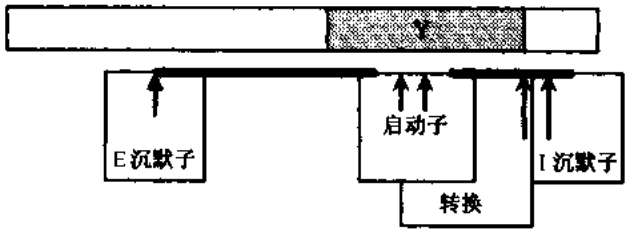


图 18－14　在活性暗盒的启动子和侧翼区域具有超敏位点。
（仿 B. Lewin：Genes Ⅵ，1997，Fig. 32－8）

SIR 蛋白作用在染色质结构上，通过抑制潜在的超敏感位点，阻断了基因的表达，DNA 复制可能使这一区域的染色质的结构发生改变，然后 SIR 蛋白才和组蛋白以及其他的核蛋白产生一定的相互作用。

(3) *MATa* 和 *MATα* 基因编码调节蛋白

α 细胞含有很多的“α－特异性蛋白”，a 细胞是不会产生此蛋白的。而 a 细胞含有一套“a－特异性蛋白”，这些蛋白也不可能由 α 细胞产生。α－特异性蛋白的基因仅在 α 细胞中转录，而同样 a－特异性蛋白的基因也只在 a 细胞中转录。

交配型座位负责细胞类型特异性基因的表达：*MAT*a 和 α 编码的蛋白激活或抑制分散在酵母基因组内的一些基因的表达。因此交配型的基因座位是细胞特化的主要控制座位。图 18－14 表示交配型座位的产物控制一组特化基因的表达。

MAT 座位的基本功能是控制外激素和受体基因以及其他和交配有关基因的表达，每种类型的座位编码一些调节蛋白。*MATα* 编码两种调节蛋白：α1 和 α2。*MATa* 编

码一种调节蛋白 a1。a 和 α 蛋白直接控制靶基因的转录，它们具有正调节和负调节的功能。它们的功能在单倍体中以及在接合后形成的二倍体中都是独立的。

在 a 细胞中，a 特异基因（*asg*）是组成形表达（图 18－15b）的，而在 α 细胞中却受到阻遏不能表达（图 18－15a）。a－特异性基因包括 a－因子结构基因和 *STE2* 基因（sterile matation）；*STE2* 编码 α－因子受体。这样 a 细胞的表型表现为能迅速识别 α 型所产生的外激素。

在 α 细胞中 α－特异性基因（α*sg*）功能是由 α1 诱导的，因此 *asg* 只在 α 细胞中表达。它们包括 α－因子结构基因，α－因子的受体基因 *STE3*。一种类型细胞外激素的表达和另一种类型受体的表达是相联系的。

单倍体特异性基因（*hsg*）包括外激素转录所需的基因和受体基因，*HO* 基因（和交配型转换有关），和 *RME* 基因（阻遏减数分裂和孢子形成的基因）。它们在两种类型的单倍体中都是组成型表达。但在 a/α 二倍体中受到阻遏（图 18－15，16），因此 *asg* 和

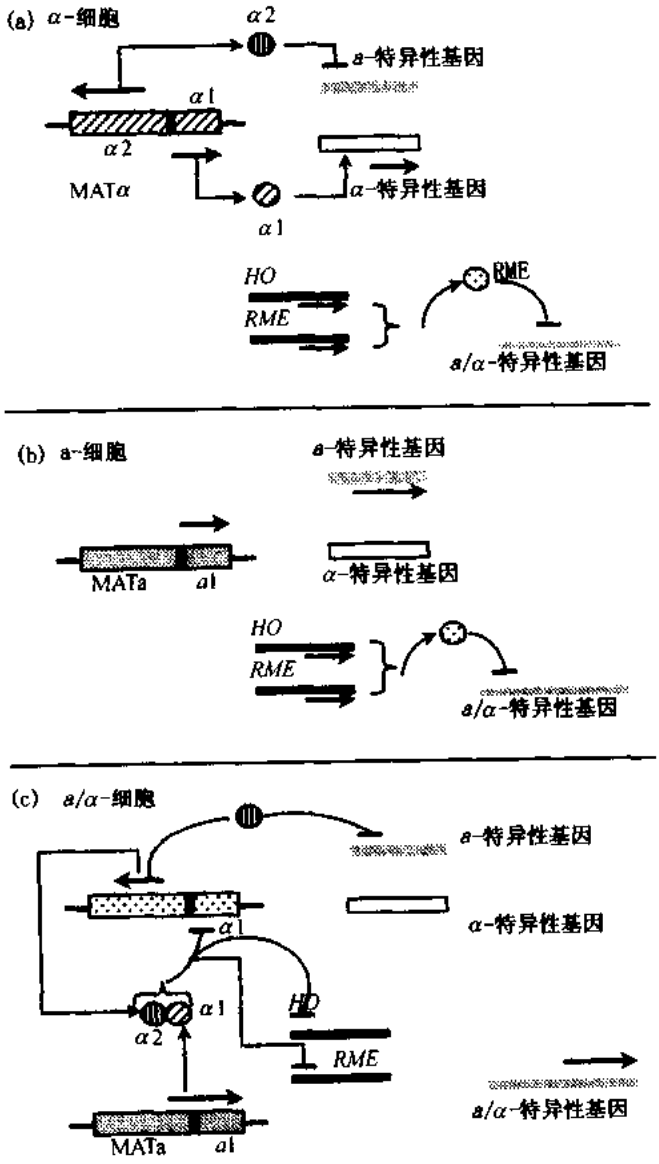


图 18－15　MAT 基因的调节功能（a）α－单倍体；（b）a－单倍体；（c）a/α 二倍体细胞。"→"转录；"↘"激活"⊣"抑制。

asg 在二倍体中都不表达。

现在的观点认为在单倍体和二倍体酵母 *MAT* 的表达决定调节功能和不同的靶基因。但 a 和 α 交配型受到不同机制的调节：①在 a 单倍体中 *asg* 是组成型表达(图 18－15b)。在 a 细胞中 MATa 产物的功能还不知道(图 18－16)，它可能是仅在二倍体细胞中起到阻遏 *hsg* 的功能；②在 α 单倍体中 *α1* 的产物激活 *asg*(图 18－15)，*asg* 的产物是产生 α 交配型所必需的。α2 的产物可以阻遏 α 细胞中负责 a 交配型的基因 (图 18－15a)。其作用是通过结合于靶基因上游操纵元件的序列来实现的；③在二倍体中 *α1* 和 *α2* 的产物相互合作来阻遏 *hsg*。它结合识别 *hsg* 的操纵元件序列。此时阻遏的靶基因和单个 α2 所识别的靶基因是不同的(图 18－15)。

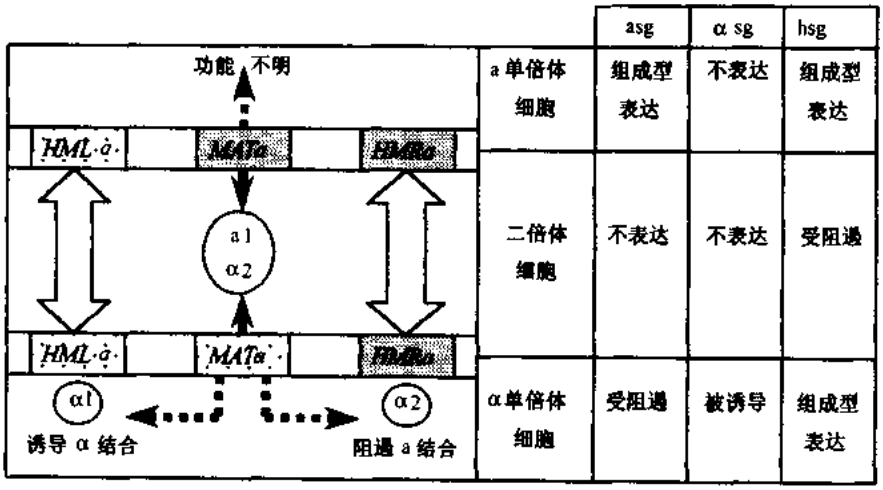


图 18－16　asg，α sg 及 hsg 在 a 细胞，α 细胞及二倍体细胞中的表达。(仿 B.Lewin：Genes Ⅵ，1997，Fig.32－6)

调节转录的 α2，a1 和 α1 蛋白能在彼此之间或和其他蛋白之间发生蛋白的相互作用。在 a 细胞，α 细胞和二倍体细胞中基因控制的模式如图 18－17 所示。

一种称为信息素和受体转录因子(pheromore and receptor transcription，**PRTF**)的蛋白对交配型是非特异的。以前被描叙为 **P 盒转录因子**。它们和上述蛋白质之间相互作用有关。PRTF 结合于一个很短的保守序列 *P box* 上，它在基因调节中的作用可能相当广泛，因为 *P box* 存在于很多位点。在有些位点中 *P box* 对基因的激活是必需的。但在另一座位上，PRTF 又是为阻遏所需。因此它的作用可能依赖于结合在 *P box* 邻接位点的其他蛋白。

asg 可能只由 PRTF 激活，在 a 单倍体细胞中 PRTF 足以保证 *asg* 的表达。在 α 单倍体细胞中 *asg* 受到 α2 蛋白和 PRTF 的共同作用而被抑制。α2 蛋白含有两个功能域，C 端功能域结合于含 32bp 保守序列的操纵基因一端的短回文序列上，但这个 DNA 片段的结合并不导致阻遏。N 端的功能区起到阻遏的作用并负责和 PRTF 接触。PRTF 的结合位点是在操纵基因中心的 *P box* 处，α2 和 PRTF 是协同结合在操纵基因上。

asg 的表达需要 α1 的激活，α1 是另一种小蛋白，长 175 aa，*asg* 转录的起始序列位于 UAS 元件中，此保守序列长 26bp，可分为两部分，第一部分是 *Pbox* 长 16 bp，这是 PRTF 结合处，相邻接的 10bp 序列形成了 α1 的结合位点称 *Qbox*。只有当 PRTF 结合于 *Pbox* 时，α1 因子才能结合于 *Qbox*。这两种蛋白都不能单独结合于它们的靶 box 上，但可以共同结合于 DNA，这是蛋白与蛋白相互作用的结果 (图

18 - 17)。

在 a 单倍体细胞中，由于缺乏 α1 蛋白，所以 *asg* 不能表达，PRTF 也不能结合 *P* box 使其激活。

α2 蛋白也可以和 a1 蛋白协作，这两个蛋白的组合识别一种不同的操纵基因。此操纵基因只有 α2 识别的回文序列，和 asg ，αsg 的特异序列都不同，比较短。在二倍体细胞中 α2/a1 复合蛋白可以结合于此模体上阻遏 hsg 的表达。

(4)转换的机制

交配型的转换描述为**基因转变**(conversion)。在此过程中受体位点(*MAT*)被转变成供体型的序列(*HML* 和 *HMR*)。*MAT* 的突变可阻止交配型转换，通过这种突变来鉴别出转座(见 25 章)所需的位点。通过 *HML* 或 *HMR* 中的缺失突变发现这个转变过程具

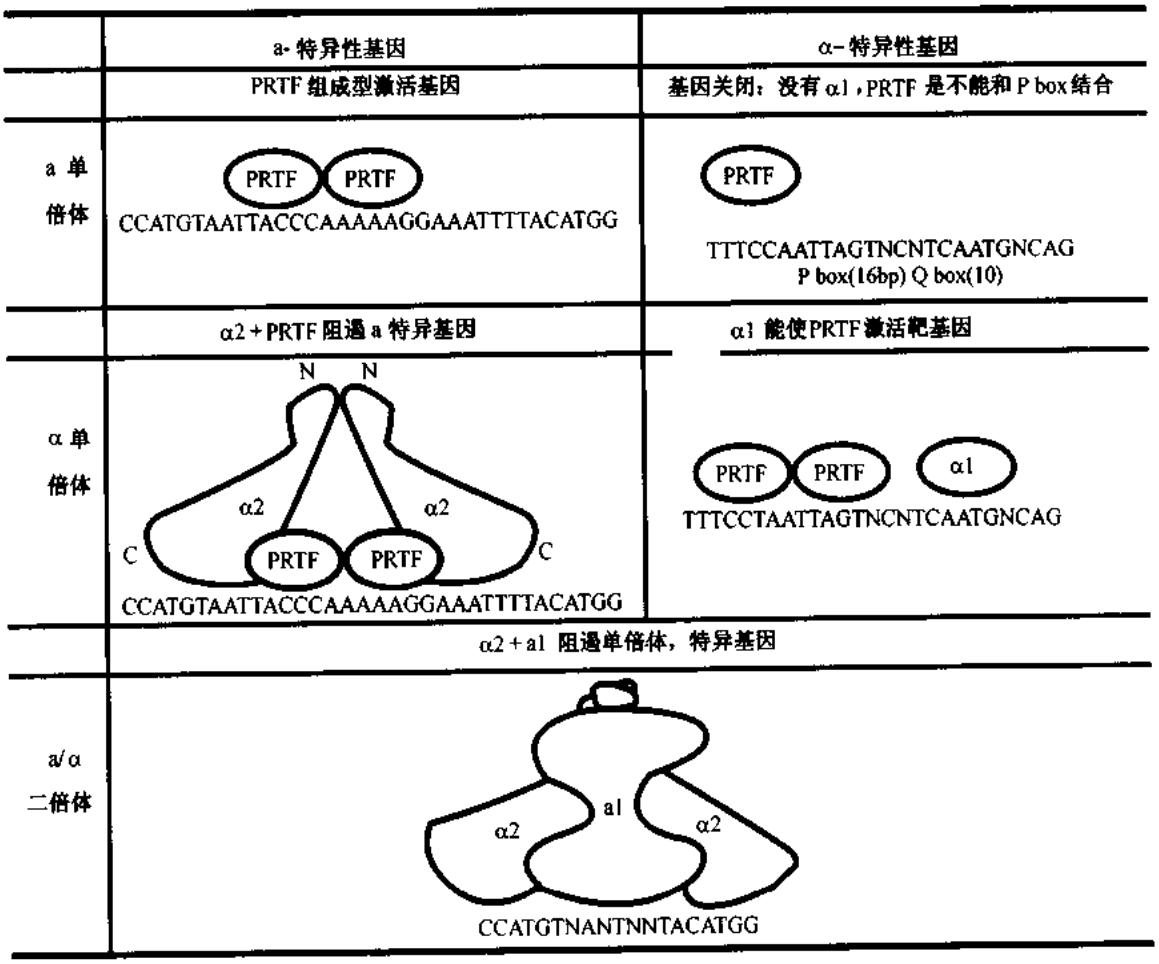


图 18 - 17　PRTF，α1，α1 和 α2 的组合激活或阻遏特殊的基因群，与细胞的交配型保持一致。
(仿 B. Lewin: Genes Ⅴ, 1994, Fig. 36 - 7)

有单向性的特点。

在 *MAT* 的 *Y* 右边界上鉴别一个突变位点，它对于转换是至关重要的。*Y* 右端的缺失并没有什么影响。但一旦缺失越过这个界线那么就不再发生转换。此边界的性质是通过和转换有关的一些点突变来分析定位的。有的点突变发生在被取代的区域中(转换之后 *MAT* 消失了)，而另一些突变正好发生在替换区的近外侧(这样就持续地妨碍转换)。所以取代区内和外侧的序列对于转换来说是必须的。

Y - *Z* 的交界处是 DNA 改变的位点，它起始转座。经历过转换的细胞群体中，约

1% ~ 3%的 MAT 座位 DNA 在此位点有双链断裂点。断裂点紧靠着交界处，并与 DNase 超敏位点相一致。

现在能提出有关转换的一系列可能的事件。DNase 超敏位点在 *Y* - *Z* 交界线上，图 18 - 18 表明剪切产生了 4 b 的单链末端。核酸酶没有附着到突变的 *MAT* 座位上，因此也不能转换。缺失分析表明。*Y* 交界处周围的 24bp 的全部序列或大部分序列对于体外剪切来说都是必要的。此识别位点对于核酸酶来说相当大，可能此识别序列仅存在于 3 个交配型暗盒上。

有理由认为使沉默暗盒不转录的机制与它们不受 HO 核酸内切酶作用的机制是相同的。结果使转换变成为单向性的。

图 18 - 19 表明由剪切，修复而产生的转换反应的机制。供体和受体之间同源配对，受体 DNA 按链断裂重组的模式进行。重组的起始需要酶的切割。这些酶的相关基因发生突变会抑制转换。

推测 *MAT* 的游离端插入 *HML* 或 *HMR* 座位且和 *Z* 区配对。*MAT* 的 *Y* 区被降解直到 *Y* 区左侧的同源区。*MAT* 的左、右两侧是可以和 *HML* 或 *HMR* 配对的。*HML* 或 *HMR* 的 *Y* 区拷贝出新的序列取代了 *MAT* 被降解的区域。配对座位再彼此分离(此过程的序列可能不同)。

与双链断裂重组模型相似，此过程是由 *MAT* 起始的，这意味着 *HML* 和 *HMR* 的最

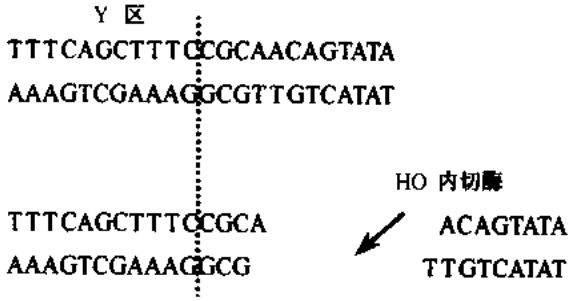


图 18 - 18　HO 内切酶在 Y 区的右侧交错剪切双链，产生带有 4 个碱基单突出的粘性末端。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32 - 9)

大作用就是作为供体座位。没有加工的机制。像复制转座一样，受体座位是经历了替换，而不是加上一段 DNA。

(5) *HO* 表达的调节

交配型的转变是由 *HO* 基因起始的，而它本身又通过相互作用的方法进行调节，使转录受到一系列的控制。①*HO* 是在交配型控制之下，因为在 *MAT*a /*MAT*α 二倍体中它是得不到合成的。可以认为 *MAT* 座位的一对等位基因表达无需交配型的转换。②*HO* 在母细胞中转录，而不在子细胞中转录。③*HO* 的转录对细胞周期作出反应，此基因仅在母细胞的 G_1 期末才得到表达。

核酸酶产生的时间说明了交配型转换和细胞谱系的关系。转换只有在分裂的产物中才能检测得到。图 18 - 20 表示 *HO* 的表达限制在 G1 期，结果在 MAT 座位复制前交配型发生转换，便两个子细胞获得了新的交配型。

控制 HO 转录的顺式作用位点位于 *HO* 基因上游的 1.5kb 处(图 18 - 21)。控制的一般模式是在很多位点中的任何一个位点上的进行阻遏，对一系列调节循环作出反

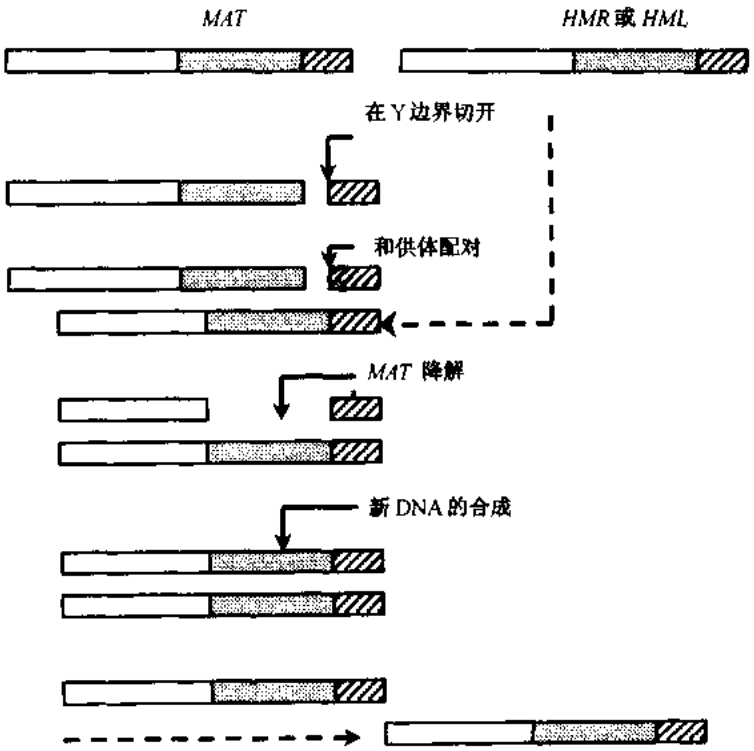


图 18－19 受体（*MHT*）座位的双链断裂，起始暗盒的替代，可能涉及到 Y 区域和供体（*HMR* 或 *HML*）座位的配对。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32－10）

应，可以阻止 *HO* 的转录。

交配型的控制类似于单倍体特异基因（*hsg*）的控制。在二倍体中转录受到 a1/α2 的阻遏。在 *HO* 基因的上游区有 10 个阻遏物的结合位点，这些位点的保守序列十分相似。对这些阻遏位点知之甚微，也不知道其中有多少对单倍体特异基因的阻遏是必要的。细胞周期控制位点是由 9 个拷贝的八核苷酸序列构成，位于 －150～－900 之间的区域。一个拷贝的保守序列能被细胞周期控制因子附着于其上。在 G_1 期末除了一个短暂的时期以外与此序列连锁的一个基因被阻遏，它的激活依赖于细胞周期调节因子 CDC28，细胞分裂时，它在起始阶段（START stage）发挥作用。

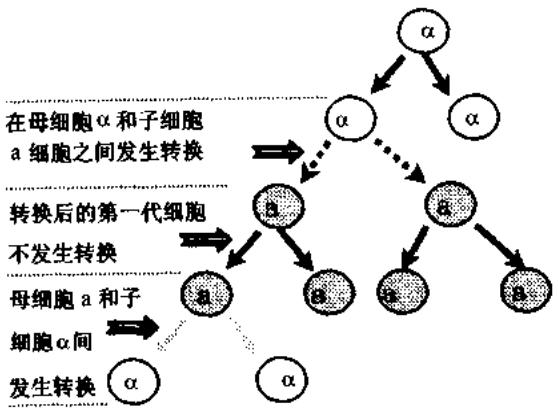


图 18－20 转换只发生在母细胞中：两个子细胞具有新的交配型。子细胞必须通过一个完整的细胞周期才能变成母细胞，此母细胞又能发生转换。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32－11）

SWI 和 *SIN* 基因之间相互作用涉及到细胞周期、母细胞和子细胞的控制。基因 *SWI1 ~ 5* 是 *HO* 转录所需要的。它们的功能是阻止 *SIN1 ~ 6* 基因产物阻遏 HO。*SWI* 基因是通过抑制转换的突变体被发现。然后通过特殊的 *SWI* 突变导致 *HO* 转录阻断而发

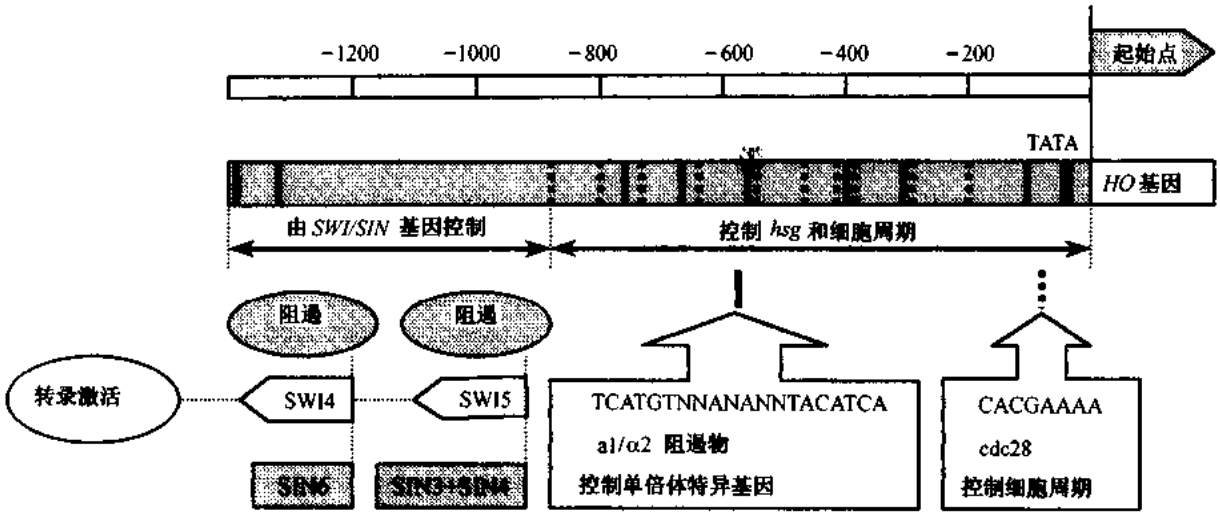


图 18 - 21　一系列的控制系统可阻遏 *HO* 基因表达，只有在所有的阻遏都失败时 *HO* 才转录。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 32 - 12）

现了 *SIN* 基因。*SIN* 基因和 *SWI* 基因之间的相互作用并未得到充分的解释，但却涉及到多种途径。

SWI5 的作用是维持母细胞中的转录。其实际功能是阻遏 *SIN 3、4* 的表达。在突变体中这些功能缺失，那么 *HO* 在母细胞和子细胞中将同样被转录。这一系统作用于上游区（- 1260 ~ - 1300）的 DNA 序列。SWI5 本身服从细胞周期调节。SWI5 表达比 HO 晚，有时在细胞周期的启动阶段前。子细胞失去了 *SWI5* 的功能。如果 *SWI5* 在第一次细胞周期被激活的话，那就不用到第二周期它就可以激活 *HO* 而导致转换。这就解释了在子细胞中转换的能力的延迟，直到它成熟成为母细胞才发生转换。（图 18 - 20）。*SWI4* 的产物以相同的形式行使功能，阻止 *SIN6* 阻遏 *HO* 基因。

HO 基因还受到 *SWI1、2、3* 的调节，这 3 个基因编码了一组蛋白，可能以复合体的形式发挥作用，很多基因的表达需要它们。因此，这种复合体可能是转录的激活物。*SIN1* 编码一种蛋白阻止 *SWI1、2、3* 的功能，而 *SIN2* 编码组蛋白 H3。此表明 *SWI1、2、3* 蛋白可使染色质的结构发生改变，此时 *HO* 和其他基因的转录是必要的，所以它们在交配型表达中所起的作用是附带的。

二、免疫球蛋白的多样性

淋巴细胞系是具有特异免疫识别功能的细胞系。按其分化，表面受体和功能的不同淋巴细胞分为 B 淋巴细胞和 T 淋巴细胞两类。T 淋巴来自于骨髓干细胞，经胸腺（Thymus）加工逐步分化而成，故称为 **T 细胞**。其功能是调节免疫应答，杀伤抗原，但不分泌抗体，故称为**细胞免疫**。B 细胞是在鸟类的法氏囊（Bursa of Fabricins）和哺乳动物的骨髓内分化成熟，不经胸腺，故称为 B 淋巴细胞。其主要功能是分泌免疫球蛋白（ig），故称为**体液免疫**。

1. 免疫球蛋白的结构和多样性

免疫球蛋白由两条**重链**(Heary chain)和两条**轻链**(Light chaim)由二硫键连接构成了 Y 型的对称结构(图 18－22)。

每条蛋白质链由两部分组成:N 端的**可变区**(variable region,V)区和 C－端的**恒定区**(constant region,C),最初 Putnam(1966)和 Edelman(1969)等通过比较不同免疫球蛋白链的氨基酸序列而鉴别出这些区域的。轻链由 213～214 个氨基酸组成,重链由 446 氨基酸组成。在重链的 CH1 和 CH2 之间有一铰连区,当抗体和抗原结合时此区可换角度,以适应不同大小的抗原决定簇。

V 区是由轻链和重链的可变区联结而形成,此区负责识别抗原。V 区的多变性反映了抗体的多样性。在可变区中又有 3～4 高变区(hypervariable region),其氨基酸序列特别多变,是与抗原分子的结合部位。在 Ig 分子中有 3 个 C 功能区 (CH1, CH2 和 CH3),不同 Ig 的 C 区显示了很高的同源性。CH1,CH2 和 CH3 每一区段承担着不同的生物学功能,因此又称为**效应区**(effector functions)如 CH2 具有活化补体的作用,CH3

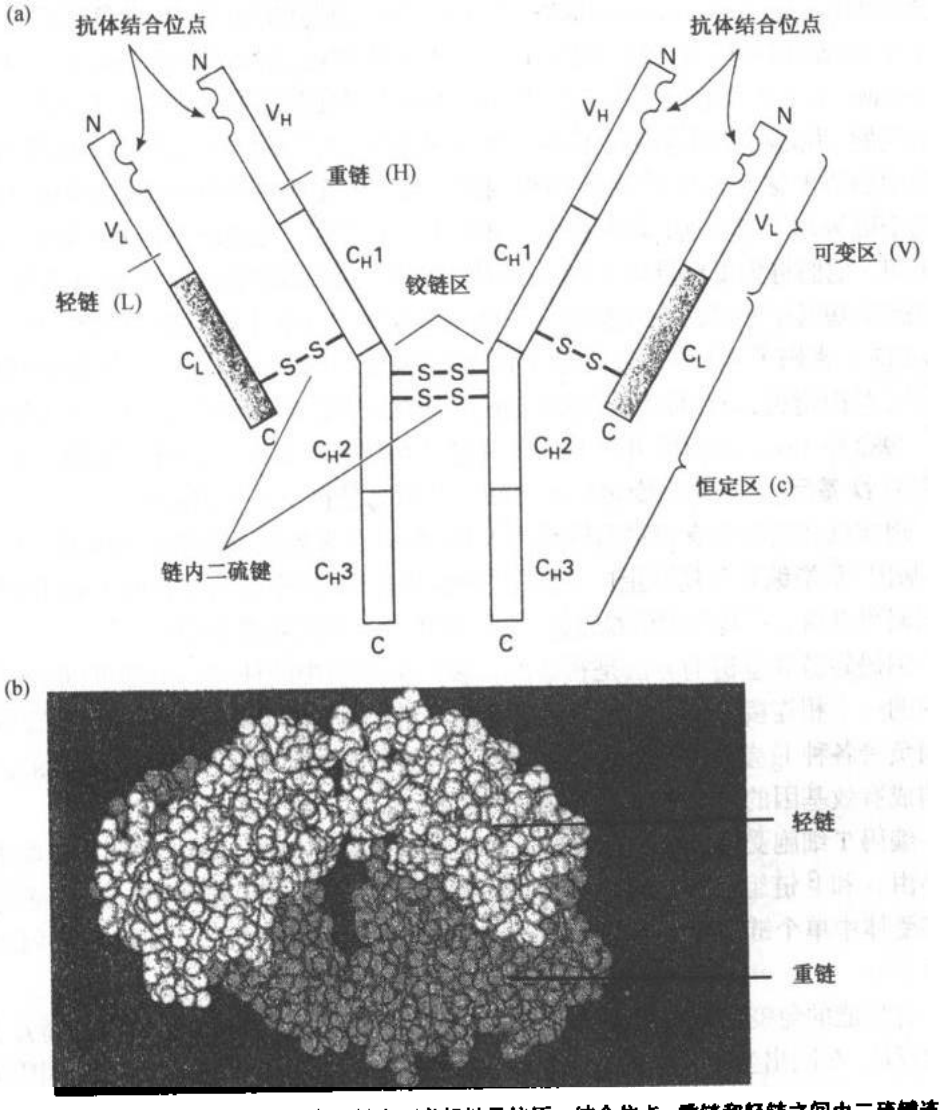


图 18－22　抗体分子。(a)图表示两条重链和两条轻链及抗原－结合位点。重链和轻链之间由二硫键连接 (b)根据 X－衍射结晶分析作出的抗体分子模型。(引自 Russell,P.J:Genetics, 3rd ed., 1992,Fig. 17－23)

具有使抗体分子粘连在单核细胞表面的功能等。

据推算哺乳动物能生成一百万种以上的抗体，而一种淋巴细胞又只能合成一种特异性的抗体，那么抗体的这种多样性是怎样产生的呢？难道有那么多的基因来编码 Ig 吗？早在 20 世纪 40 年代著名的理论化学家 L.Paulinng 就提出了**指令学说**来解释抗原的多样性，他提出抗原的多样性可能是由于蛋白质二级结构的改变所致，而一级结构可能只有一种或少数几种，所以相应的基因也不多。后来人们发出蛋白的一级结构决定了其二级及三级结构，因此这种学说不能成立。

1957 年 F.M.Burnet 提出了**克隆选择学说**，主要论点是：①一种浆细胞只能产生一种抗体；②抗原起到了选择某一淋巴细胞克隆的作用。很多的实验都支持这一观点，但仍不能解释哺乳动物总的基因数不过 10 万，怎么可能有千万个以上的基因来编码不同的抗体。

对克隆选择的解释有三种：①在生殖细胞中就已存在大量的 V 区基因。此仍不能解释 Ig 的多样性和基因数有限的矛盾；②Burnet 又提出了体细胞突变的假说来解释多样性。但若是体细胞突变所造成的话，这样高的突变率也是难以解释的；③在 20 世纪 60 年代美国加州理工学院的 W.T.Dreyer 和 J.C.Bennett 修改了**胚系学说**，提出了**体细胞重组**(somatic recobination)**学说**，认为在浆细胞成熟的过程中，染色体发生了重排，多个 V 区基因中的一个和 C 基因组合，产生一条 DNA。1976 年 Mozumi,N. 和利根川进(Tonegawa, S.)用实验证实了 IgG 基因中 DNA 的体细胞重排，从根本上解决了 Ig 多样性的问题，为此利根川进荣获了 1987 年诺贝尔奖。他们用分子杂交的方法发现在胎鼠细胞中编码 V 区和 C 区的 DNA 序列相隔较远。但在成熟的抗体细胞中(淋巴瘤细胞)这两个区域却紧密相连；表明在淋巴细胞分化过程中，免疫球蛋白基因曾发生过体细胞重组。他们将免疫球蛋白 DNA 在细菌中克隆，在胎鼠细胞的一个 DNA 克隆中，测定 λ 轻链 V 基因序列时，意外发现 V 基因只编码 V 区 108 个氨基酸中的 95 个，最后的 39bp(13 个密码子)消失了，那么余下的 13 氨基酸的编码序列在什么地方呢？他们发现是在 C 基因附近，现在称为连接(Joining,J)区附近(连接 V 和 C 基因的 DNA 序列)。

1982 年 Hood 和利根川进等又在 H 链的编码区中发现了**多样性基因** (Diversity)，简称为 ***D* 基因**，它编码高变区的氨基酸。其位置是在 V、J 基因之间。

现在已经搞清楚有很多基因编码 V 区，而只有少数几个基因编码 C 区。以上所说的“基因”是指编码免疫球蛋白多肽链(重链或长键)的不连续部分的 DNA 序列。V 基因编码可变区 ，C 基因编码稳定区，它们都作为一种独立表达的单位。

编码轻链和重链的方法是相似的：多个 V 基因中的任何一个都可以和少数 C 基因中的一个相连接。体细胞重组是发生在 B 淋巴细胞中(表 18－3)。大量的不同的 V 基因负责各种 Ig 差异的主要部分。然而不是所有的差异都是由基因组编码的。有的是在构成有效基因的过程中发生改变而产生的。

编码 T 细胞受体蛋白链的有效基因的形成也是相似的。在 T 细胞中有两类受体，一类由 α 和 β 链组成，另一类由 γ 和 δ 链组成。与编码 Ig 的基因相似，这些基因编码 T 细胞受体中单个链，构成分离的部分，包括 V 区和 C 区，这两个区集合在活化的 T 细胞中。

有功能的免疫球蛋白或 T－细胞受体基因存在着一种改变，以对表型特点(抗原)作出反应。在刚出生时，生物并不具有产生特殊抗体或 T 细胞受体的有效基因。而具有大量的 V 基因和少量的 C 基因。随后由这些基因构建有活性的基因，从而能合成抗件及受体，以对抗原作出反应。克隆选择学说要求这种 DNA 的重排发生在抗原进入体内

之前，然后抗原选择了产生相应抗体的细胞。整个的过程都发生体细胞中，且不影响生殖细胞。因此对抗原的特殊反应并不遗传给后代。

V、C 基因间的重组产生了有功能的位点，这发生在未成熟的淋巴细胞中（表 18－3）。一个 B 淋巴细胞通常只有一个轻链的基因和一个重链基因重排；而在 T 淋巴细胞中一个 α 基因和一个 β 基因产生重排，或者一个 δ 基因和一个 γ 基因产生重排。*V*、*C* 基因特殊的连接决定了抗体或 T 细胞受体的产生。

Ig 轻链有两家族：*K* 和 λ，而重链只有一个家族（表 18－4），每家族位于不同的染色体上，由自己的一套 *V* 基因和 *C* 基因构成。在任何动物中每家族的 *V* 基因和 *C* 基因都隔开一段距离，这种排列的方式称为胚系（germ line）模式，在生殖细胞中和除了免疫系统外的所有体细胞中都是按此排列。

但在表达某种抗体的细胞中，每条链是由单个完整的基因编码的。V 基因和 C 基因重组产生了由外显子构成的有活性的基因，外显子的位置与蛋白质的功能域保持一致。内含子通常被剪切掉。在一个家族中 *V* 基因和 *C* 基因的数目随着物种的不同而异。表 18－4 表示人和小鼠每个 Ig 家族中的成员。最典型是有几百个 *V* 基因，但要注意人和小鼠之间的差异在于小鼠 λV 基因丢失了很多。

在每个家族中功能性基因原则上是相似的，但 *V* 基因和 *C* 基因组成的调节存在差异，且二者之间要进行重组。

表 18－3　B 细胞的分化过程

分化时期	分化途径及特点	组织
抗原非依赖期或克隆扩增期	干细胞 ↓ 前 B 细胞　完成选择性重排，IgM 产生 ↓ 未成熟 B 细胞　膜表面有 IgM ↓ 成熟的 B 细胞　膜表面有 IgM，IgG，IgA 等作为抗原受体 ↓	骨髓
抗原依赖期或克隆选择期	浆细胞　膜表面 Ig 消失（受体功能消失）胞质内合成大量 Ig，并分泌	外围淋巴细胞

表 18－4　不同的 Ig 家族的 V.D.J.C 基因数

家族	*V*		*D*		*J*		*C*	
	人	小鼠	人	小鼠	人	小鼠	人	小鼠
Lλ	<300	2			76	4	>6	4
LK	<300	~1000			5	5	1	1
H	~300	>1000	~30	12	4	4	9	8

2. 轻链基因和重链基因各自的重组

（1）轻链基因的重组

在小鼠胚系 DNA 中，有很多的基因片段编码 κ 轻链（图 18－23），其中含有 3 种片段（19～22）：① *L*－V_κ 由前导序列和 V_κ 组成。V_κ 编码轻链可变区的大部分氨基酸。*L* 编码信号肽（见第 16 章），这是 Ig 分泌所必须的。但并不参与抗体分子的功能。②C_κ 片

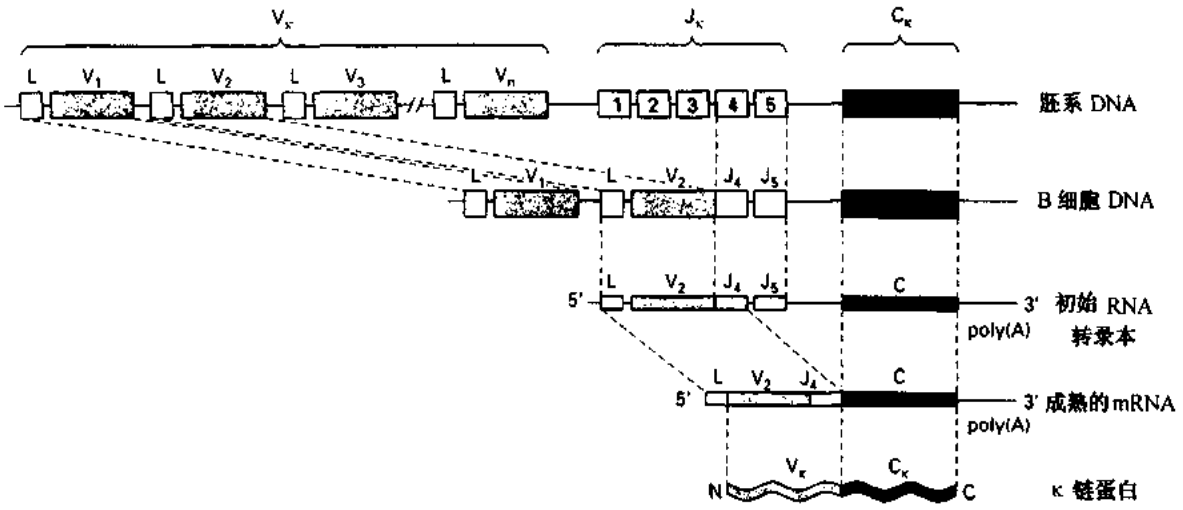


图 18－23　小鼠轻链基因在发育中通过 V、J 和 C 基因片断的重组产生了 DNA 的重排。此图表示很多重排中的一种情况。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 17－24)

段：编码 κ 轻链的 C 功能区，③ J_κ 片段：连接片段在有功能的 κ 轻链基因的产物中用来连接 V_κ 和 C_κ。

在前 B 细胞(pre－B cell)中，$L-V_\kappa$ 和 C_κ 片段依次排列在染色体上。每个 $L-V_\kappa$ 片段长约 400bp，它们串联排列在染色体上，长约 7kb。J_κ 片段长约 30bp，串联排列占据约 20kb。分散在 J_κ 和单个 C_κ 中的内含子，每个长约 2～4kb，当 B 细胞发育时，一个特殊的 $L-V_\kappa$ 片段和其中一个 J_κ 连接，再和 C_K 连接(图 18－23)，然后转录，切除内含子，成为成熟的 mRNA，它由 $L-V_{\kappa 2}J_{\kappa 3}C_\kappa$ 组成，翻译后进行分泌，前导肽被切除，此时 B 细胞中产生了 Ig 的 κ 轻链。

在小鼠中约有 1000 个 $L-V_\kappa$ 基因片段，4 个有功能的 J_κ 片段和 1 个 C_κ 基因片段。这样就可能产生 $1000\times4=4000$ 个不同的 κ 链。同样的机制也存在于小鼠 λ 轻链的基因中。但 $L-V_K$ 基因片段只有 2 个，而有 4 个 Cλ 基因。每个 Cλ 都带有它自己的 Jλ 基因片段，这样产生 λ 链的可变区要比 κ 链少得多。V_k 和 J_k 基因的不同的连接产生了 κ 链的多样化。V_κ 基因的 3′端和 J_K 基因的 5′端具有特异的识别序列，作为 V_κ 和 J_κ 的重组信号。这些识别序列估计是由一种尚未鉴别出的酶加以识别并催化其重组。在 $V_\kappa J_\kappa$ 连接时 V_κ 和 J_κ 丢失了几对核苷酸，这样在连接的地方产生了有重要意义的多样化。这样 κ 轻链的多样性是由于下列因素而产生的：①多种 V_κ 基因片段的存在；②4 种 J 基因片段的不同；③在 V_K-J_κ 连接时不同数量碱基对的缺失。

(2)重链基因的重组

Ig 重链基因也是由 V_H、J_H 和 C_H 片段编码的。但比轻链还多了个 D(diversity)基因片段，它位于 V_H 和 J_H 片段之间(图 18－24)。图 18－34 表明在小鼠的胚系中 $L-V_H$ 串联排列，然后是一组 12 个 D 片段，再接着是 4 个 J_H 片段。在恒定区基因是由一组基因组成，在小鼠中此基因簇依次为 μ、δ、γ_3、γ_1、γ_{2b}、γ_{2a}、ε、α 八个基因。分别编码重链 IgM, IgD, IgG, IgE 和 IgA 的恒定区，当与 L 链基因重排时，抗体的多样性除了 V、D、J、C 的不同连接外，还依赖于 L 链和 H 链的连接。

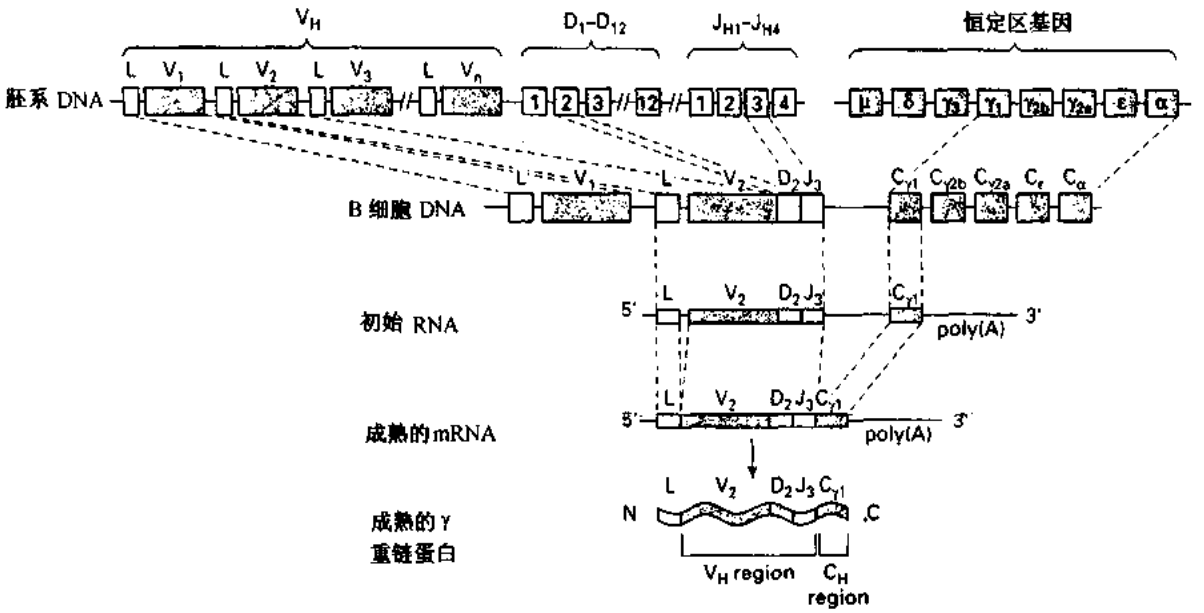


图 18－24　小鼠重链基因在发育中通过 V、J 和 C 基因片断的重组产生了 DNA 的重排。根据 C_H 片断的使用，抗体分了可能是 IgM、IgD、IgG、IgE 和 IgA. 此图表示一种 IgG 的组合，这是很多种重排中的一种情况。
（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 17－25）

3. *V*－*C* 基因间的重组产生缺失及重排

轻链和重链的组装涉及到同一机制，在所有的胚系片段的分界线处都有相同的保守序列，这些序列参与连接反应。每一个一致序列都含有**七聚体**（heptamer）和**九聚体**（nonamer），中间被 12bp 或 23bp 的间隔序列所分开；图 18－25 表示小鼠 Ig 座位。在 κ 座位上，每个 V_κ 基因后面都很随着一个带有 12bp 间隔序列的一致序列。每个 J_κ 基因的前面也有一个带有 23bp 间隔序列的一致序列。*V* 和 *J* 的一致序列是反向排列的。λ 座位的情况也与此相似，但排列的序列相反。在 H 链中 V_H 和 *D* 之间有两个一致序列，在 *D* 和 *J*－C_H 之间的 2 个一致序列，以 *D* 为对称中心方向相反，对称排列。

在一致序列的成分之间总是间隔着 1 个或 2 个双螺旋转角，这一特点可能反映了重组反应中的几何关系。例如重组蛋白可能要从一侧接近 DNA，就像 RNA 聚合酶和阻遏蛋白接近识别启动子和操纵子类似。

用“重组”来描述 Ig 基因成分的连接并不意味着此反应涉及同源序列之间的交互

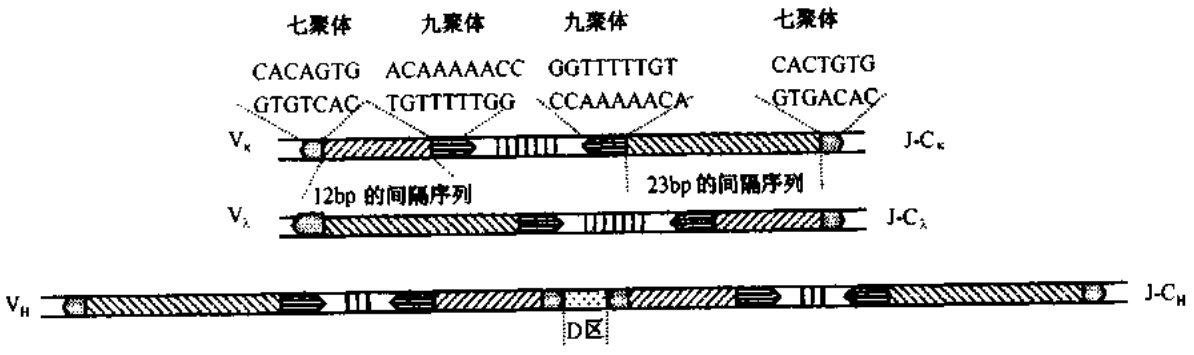


图 18－25　重组位点的保守序列反向或成对排列。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 33－12）

重组。存在着一种 DNA 序列的物理重排，涉及到断裂和重接，但机制和同源重组不同。图 18－26 是以 κ 轻链为例来说明这种重组的一般特点。重链座位的重组反应也是如此，只不过发生两次连接(先 *D－J*，后 *V－DJ*)而已。

断裂和重接是分开发生的。在七聚体靠近编码单位的一端双链断裂，将 *V* 基因和

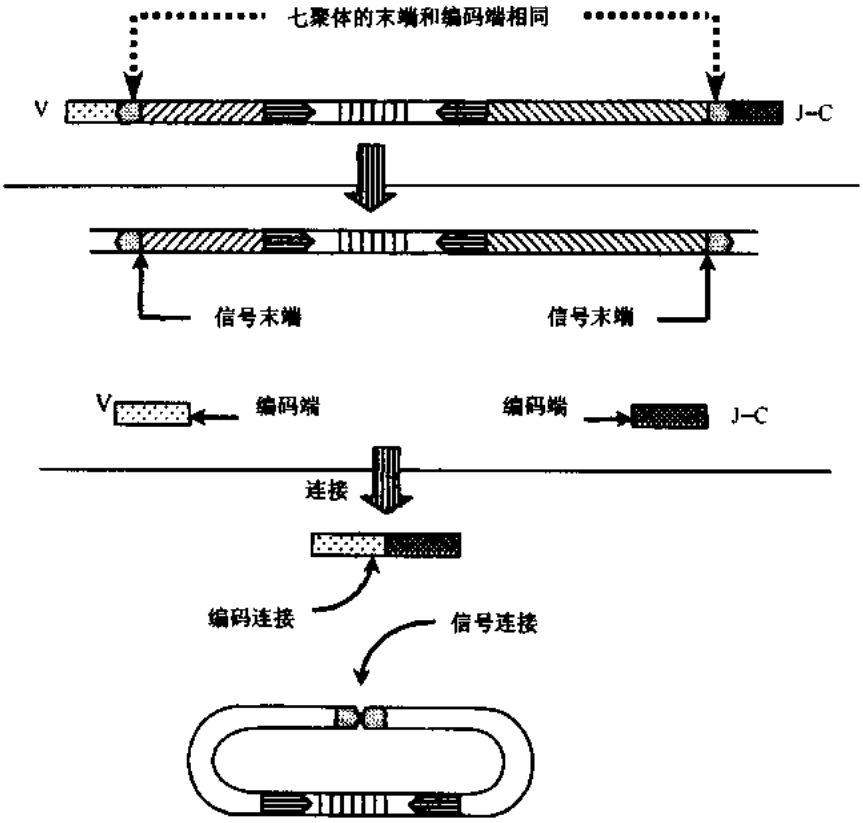


图 18－26　保守序列的断裂和重接产生 Ig 基因。(仿 B.Lewin:Genes Ⅵ,1997,Fig.33－13)

J－C 基因之间的完整片段释放出来，这个片段被剪切的末端被称为**信号端** (signal ends)V 和 J－C 座位的剪切末端称为**编码端**(codig ends)。两个编码端共价连接形成**编码连接**(coding joint)，将 V 和 J 片段连接起来，如果两个信号端也相连接，那这个被切下来的片段也会形成环状分子。

这种连接反应导致序列的改变，影响到 L 链上 V－J 连接处或 H 链上的 V－D 及 D－J 连接处的氨基酸编码。这种改变是酶在起作用，使 DNA 断裂和重接。在以上的重组过程中会产生碱基的丢失或插入，图 18－27 表示碱基对增加的机制。缺失也发生在 VL－JL 连接中，但插入在这种连接中是较罕见的。

在重组机制的研究中主要的障碍是缺乏体外系统。人们试图从 B 细胸和 T 细胞提出液中分离特异重组的 V 和 C 基因，但至今都失败了。一种间接的方法是将胚系序列的特殊重组导入淋巴细胞，再研究重组的产物。

人们在此反应的起始阶段从带有 SCID(severe combined immunodeficent，重度联合免疫缺损)突变的小鼠淋巴细胞中分离了中间产物，并加以鉴定。*SCID* 突变的原因尚不清楚，但结果导致了 Ig 的活性及 TCR(T cell receptor)重组大大降低。*SCID* 小鼠积聚了编码端双链断裂的分子，这种突变可能影响到连接反应的某些方面。

此编码端有一种稀有的结构：改变的发夹回，即一条链的 3′端和另一条链的 5′端

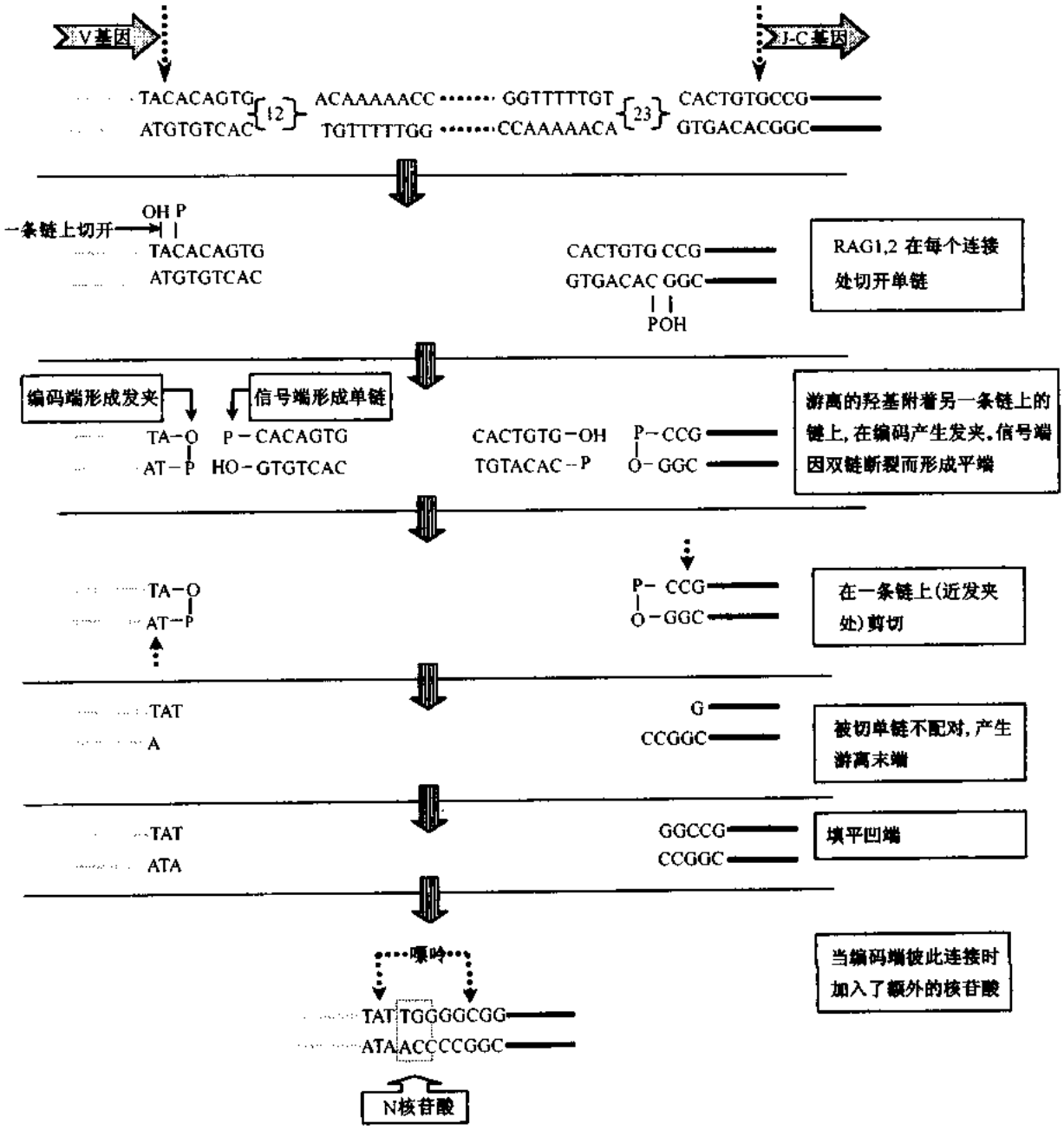


图 18-27　在编码端断裂和重接时加入了噻吟和任意的核苷酸。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 33-34)

连接。若在紧靠发夹回处单链发生断裂,则在末端出现不配对反应,产生了单链互补合成使单链凸起。互补合成使单链游离的编码端变成了延伸的双链。这个反应解释了在编码端怎样导入了 **P 核苷酸**(P nuclectices),它们构成了几个新的额外碱基对,恢复了编码端,但和原来的方向相反。

有的额外碱基也可能是插入的,位于编码端之间,表面上带有随机序列,它们被称为 **N 核苷酸**(N nucleotide),它们的插入是在连接过程中通过脱氧核苷转移酶(已知是淋巴结的一种活性成分)的活性在 3′游离编码端产生的。

这几种机制共同作用的结果使得编码连接可以产生与 *V*, *D* 和 *J* 区编码端直接

连接不同的序列。

在连接中，序列的变化引起这个位点氨基酸编码的重大改变。第 96 位点插入的氨基酸就是通过 *V* - *J* 连接反应产生的。它形成了抗体结合位点的部分，且和 L 链与 H 链之间的连接也有关系，所以与靶抗原接触位点产生极大的多样性。

碱基对数目的改变影响到阅读框，连接过程产生随机序列和阅读框有关。通过连接可能只有 1 ~ 3 连接序列保留了合适的读框。例如，若 *V* - *J* 连接，*J* 片段有**相位差**，它被转录成错误的阅读框。这样导致基因转变，因此它的表达被错误的阅读框中的无义密码子所终止。可以认为畸变基因的形成是作为增加多样性所付出的一种代价。

H 链的 *D* 片段在连接反应中也同样产生很大的多样性，同样影响到阅读框。通过连接 *J*, *C* 与前面的 *V* 基因如产生相位差，形成了非生产基因。

V 和 *J* - *C* 座位原来是同方向的，在每一个一致序列上剪切将此一致序列作为线性片段释放，若信号端被连接了，它就变成一个环状分子(图 18 - 26)。释放一个被切除的环而产生的缺失是 Ig 和 TcR(T cell receptor)位点主要的重组模型。但有时 V 基因在染色体上和 *J* - *C* 位点是反方向的。在这种情况下，断裂和重组使间插序列方向相反以取代缺失。缺失和倒位与顺向及反向重复之间的同源重组情况相同(见 22 章)。但还有一些附带的条件带有反向 *V* 基因的重组需要信号端的连接才能产生，否则在这个位点是一个断裂。倒位发生在 TcR 重链中，有时也发生在 L_κ 链位点。

与 *V* - (*D*) - *J* 连接有关的一系列蛋白都已被描述，其中包括与七聚体，九聚体结合蛋白及具有内切酶活性的蛋白，但有直接证据的仅 ***RAG***(recombination active gene)基因。*RAG1* 和 *RAG2* 是两个基因，在染色体上相距 10 kb，将它们转染到成纤维细胞中可导致适当的 DNA 底物进行 *V*(*D*)*J* 连接反应。这表明它们的产物能直接形成重组或者能激活一种辅酶。缺失 *RAG1* 或 *RAG2* 的小鼠其 Ig 或 TcR 就不能重组。结果产生未成熟的 B 和 T 淋巴细胞。

V, *C* 基因的连接与它们的活性之间有什么联系呢？未重排的V基因是没有活性的，当 *V* 基因和 C_κ 基因连接后导致转录。由于 *V* 基因上游序列并没有通过连接发生改变，在非重排基因，无效重排基因和有效重排基因中的启动子必定是相同的。位于每个 *V* 基因上都有启动子，但是无活性。它必须通过移位到 *C* 区才被激活。这个作用依赖于下游序列。下游序列到底起到了什么作用呢？一个增强子位于 *C* 基因内或在其下游，它可以激活 *V* 基因的启动子。这个增强子是组织特异性的，它只在 *B* 细胞中才有活性。图 10 - 28 表明 *V* 基因一旦移到增强子的作用范围内，其启动子便立即被激活。

4. 等位排斥

每个 B 细胞只表达一种类型的轻链和重链，因在特定的的淋巴细胞内各种类型的链只有单个的有效重排，产生一个轻链基因和一个重链基因。因为重排只涉及到一对同源染色体中的一条染色体上的基因，在另一条染色体上的等位基因并不表达，这种现象就称为**等位排斥**(allelic exclusion)。等位排斥的发生使得分析体细胞重组变得更为复杂。问题是在一条同源染色体发生了重排时在另一条同源染色体上也测得相应的等位基因。因此我们可能不得不去分析一对同源染色体的有关情况。

重排激活基因通常所显示的模式可以用已重排的 *V* 和 *C* 座位的缺失来加以说明。在活化的细胞中有两种类型的基因组成：①活性基因的探针可以揭示基因组成的重排

模式和胚系模式(非重排模式),推测连接只发生在一条染色体上,另一条染色体是保持完整的。②发现两种不同的重排模式,表明染色体重排是独立进行的。有时经重组 *V* 和 *C* 基因之间的 DNA 在细胞系中完全消失了。

当两条染色体都发生过重排时,通常其中只有一条是**有效重排** (productive rearrangement),可产生有功能的基因;而另一条是无效重排 (nonproductive rearrangement),这种重排可能有多种多样,但无论怎样它们的基因序列不能表达成为 Ig 链(它可能是重排不完全,例如发生 *D* - *J* 连接,但没有随之发生 *V* - *D* 连接,或者是异常的连接,虽然连接是完全的,但不产生一个能编码有功能蛋白的基因)。有效和无效重排同时存在,表明存在一种反馈调节控制着重组。图 18 - 29 是一个模型的轮廓,假设在每个细胞中开始时两个座位都是处在未重排的胚系模式,以 Ig° 来表示。后来这两个位点或是发生重排,产生一个有效的基因 Ig^{+},或产生无效基因 Ig^{-}。

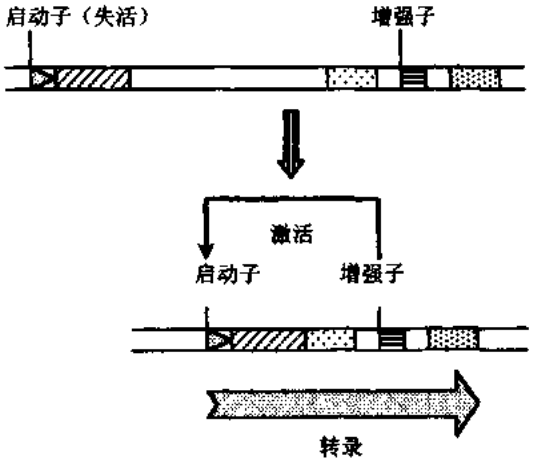


图 18 - 28 V 基因移到 C 基因的附近,其启动子被 C 基因的增强子激活。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 33 - 15)

如果重排是无效的,那么将产生一个 Ig°/Ig^{-} 的细胞,这个细胞并不妨碍未发生重组的胚系型等位基因进一步发生重排。如果第一次重排是有效的,那么细胞的基因型为 Ig^{+}/Ig^{-} 细胞。在某些情况下 Ig^{-}/Ig^{-} 细胞可试着再进行重排。有时观察到 DNA 进行连续重排的模式。此模式的关键是细胞保持 *V*, *C* 基因重组的倾向,直到达到有效重排为止。等位排斥是有效重排抑制的结果。这个机制在体内的研究是通过转基因鼠来进行的。这种小鼠的胚系有一个重排的 *Ig* 基因。转基因在 B 的细胞中的表达抑制了内源性基因的重排。

等位排斥对于 H 和 L 链位点来说是独立的。H 链基因通常先重排。L 链等位排斥适用于两个家族(细胞可能有活性的 L_{κ} 链或 L_{λ} 链)。*K* 基因先发生重排。若两条 K 链都没有试排成功的话,就再试重排 λ 链。

在 *H*, κ 和 λ 座位的重组反应中涉及到相同的保守序列和相同的 *V*(*D*)*J* 重组。三个座位仍按一套次序重排,那么为什么 *H* 链要在 *L* 链前优先重排?这可能由转录而引起的。也许这些位点是在不同的时间和酶接触,若转录发生在重排之前,κ 链为什么要优先于 λ 链进行重排呢?可能作为转录的结果。也许是这些位点在不同的时间和酶接触,转录发生在重排之前,当然这些产物没有编码功能。转录可能会改变染色质的结构,使和重组有关的任何序列能被酶所识别。

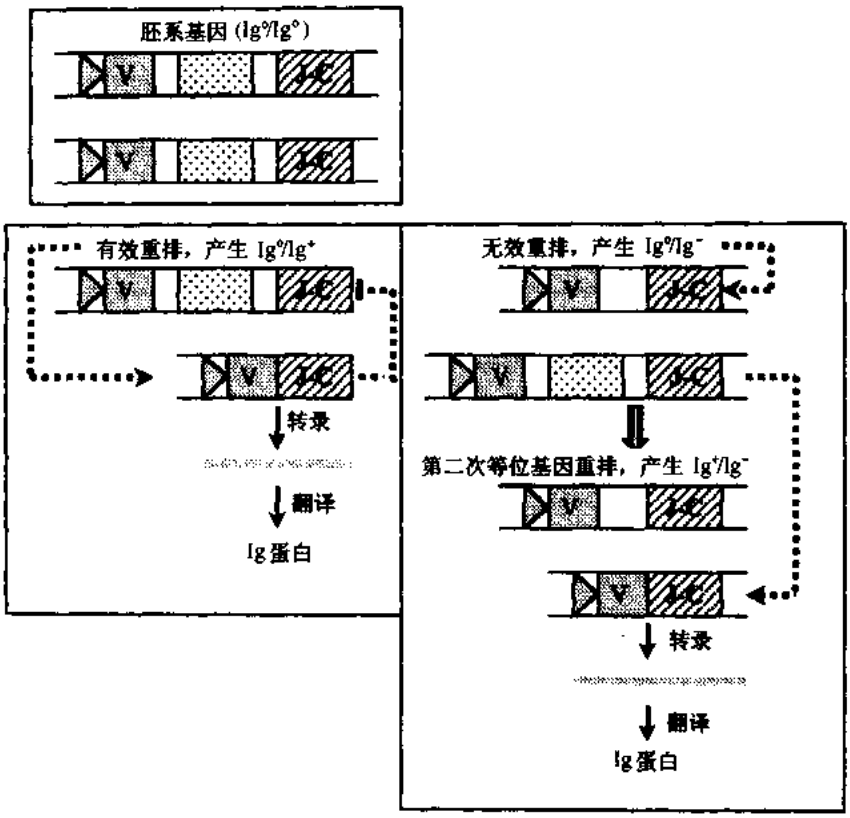


图 18－29　有效重排产生了有活性的的 H 和 L 链，并阻遏了等位基因发生同样的重排。
（参考 B. Lewin: Genes Ⅵ, 1997, Fig. 33－16）

5. 类别转换

Ig 的**类别**(class)是由 C_H 的类型决定的，表 18－5 表明 5 种 Ig 类别，现在尚不清楚每种类别的详细的功能。IgM(任何 B 细胞都产生的第一类 Ig)和 IgG(最普遍的 Ig)具有一定的活化补体的能力，导致侵入细胞的破坏。IgA 在分泌液中(如唾液)，而 IgE 和变态反应有关，且可抵御寄生虫。

所有的淋巴细胞开始时是作为一种未成熟细胞产生，承担 IgM 合成(表18－5)。C_H 基因是按胚系类型排列(图 18－30)，V－D－J 连接反应触动 C_μ 基因的表达。一个淋巴细胞在任何时候一般只产生一种类别的 Ig，但在细胞系中这种类别可以改变。这种表达的改变称类别转换(classswitching)，它是通过替换被表达的 C_H 区域来完成的。环境的作用能刺激这种转换。例如生长因子 TGFβ 能导致 Cμ 转换成 Cα。

转换只涉及 C_H 基因，相同的 V_H 基因仍持续表达。这样 V_H 基因在与一个以上的 C_H 基因重组中可以持续表达。整个的细胞系相同的轻链可以持续表达。因此类别转换可使效应反应的类型发生改变(由 V 区域介导)。

C_H 基因表达的改变有两种途径。主要的途径是 DNA 的进一步重组；另一途径是发生在 RNA 加工的水平上。但一般这种情况只涉及 C_H 区域 C 末端的改变，而不影响其类别。

表 18 - 5 Ig的类型和功能是取决于 H 链

类型	IgM	IgD	IgG	IgA	IgE
H 链	μ	δ	ν	α	ϵ
结构	$(\mu_2L_2)_5J$	δ_2L_2	ν_2L_2	α_2L_2	ϵ_2L_2
百分比	5%	1%	80%	14%	< 1%
作用功能	激活补体	耐受性发育	活化补体	存在于分泌中	变态反应

B. Lewin: Genes Ⅴ, 1994, Table 37.2

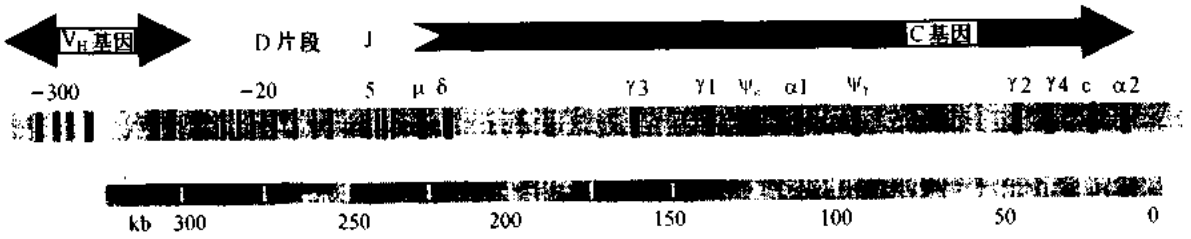


图 18 - 30 人类 IgH 链基因簇所含的全部信息。(引自 B. Lewin: Genes Ⅵ, 1997, Fig. 33 - 10)

表达 C_H 区域后部基因的细胞出现 C_μ 基因的缺失，而且位于此表达基因前部的其他基因也缺失了。类别转换是通过重组，将一个新的 C_H 基因带到与表达的 $V-D-J$ 单位毗邻的位置上。转换的 $V-D-J-C_H$ 单位的序列有一个转换位点，位于 C_H 基因的上游。转换位点称为 **S 区**(S regions)。图 18 - 31 表明两次连续的转换。在第一次转换中，C_μ 的表达后 $C_{\gamma 1}$ 连续表达，此 $C_{\gamma 1}$ 基因是通过 S_μ 和 $S_{\gamma 1}$ 位点之间的重组移到表达区域的。$S_{\gamma 1}$ 和 S_μ 之间的 DNA 缺失，S_μ 位点位于 $V-D-J$ 和 C_μ 基因之间，$S_{\gamma 1}$ 位点位于 $C_{\gamma 1}$ 基因的上游，$V-D-J$ 和 $C_{\gamma 1}$ 基因之间的区域在 RNA 加工时作为内含子被切除了。

线形缺失模型强加给重链 C_H 基因座位一种限制：一旦产生类别转换，在 C_μ 基因和新的 C_H 基因之间的任何 C_H 基因都不能表达，在图 18 - 31 例子中，$C_{\gamma 1}$ 的细胞将导致 $C_{\gamma 3}$ 不能表达，因为它在重组中缺失。

按照这个模型在已表达基因下游的任何 C_H 的基因都能进行另一次的类别转换。图 18 - 31 表明的第二次转换使 C_α 表达，它是通过 S_α 和转换区 $S_{\mu,\gamma 1}$ 之间的重组来完成的，而 $S_{\mu,\gamma 1}$ 是通过第一次转换产生的。

我们假定所有的 C_H 基因编码序列的上游都有 S 区。但我们还不知道它们是随机的还是非随机地获得后面的 C_H 基因。我们很想知道 IgM 是否能直接地转换成任何其他的类别。

我们知道转换位点并不是惟一被确定的，因为表达相同 C_H 基因的不同证明在不同位点发生了重组。当足够多的 S 位点被测序时，我们才能确定转换能够发生的限制区域。同时我们能细察此区与共同特点有关的序列。

S 区位于 C_H 基因上游 2Kb 左右，S 区的长度变化很大，从 1 ~ 10kb，含有几种短的同源重复序列。重组反应并非是位点特异性的，但通常发生在保守的重复元件的附近。关于反应的分子基础知之甚微，但在转换位点之间释放出被切除 DNA 分子。由于 S 区域位于 C_H 基因编码区前面的内含子中，因此转换并不改变翻译的阅读框。

IgM 期其合成开始于淋巴细胞，IgM 发育分成两部分，以 μ 恒定区的不同转录本进行合成。(1) 当干细胞分化成前 - B 淋巴细胞时，相应的轻链被合成了，IgM 分子

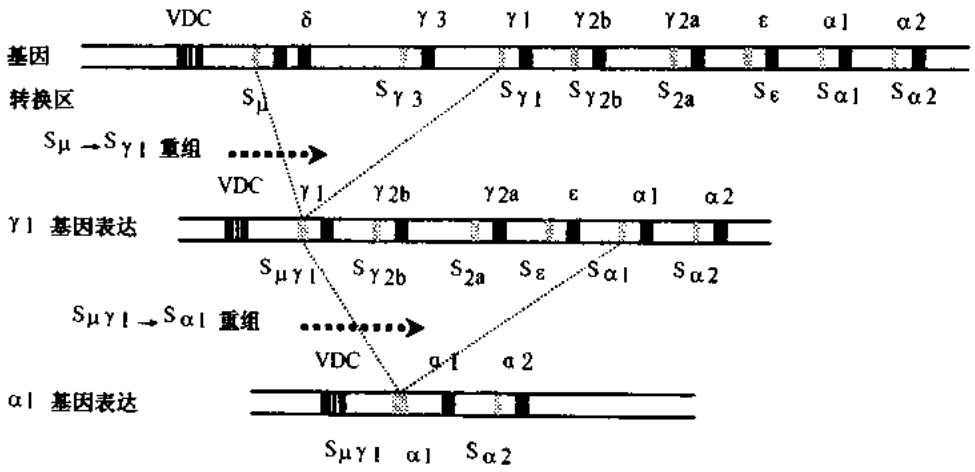


图 18－31　重链的类别转换可能由于 S 区之间的重组和重组 S 位点间的缺失所致。

(L2μ2)存在于细胞的表面。IgM 的这种形式包含着恒定区的 μm 译本(m 表示 IgM 在膜上的位置)。这种膜定位可能与抗原初始识别反应中起始细胞扩增有关。(2)当 B 淋巴细胞进一步分化为浆细胞时，恒定区的 μs 译本被表达了，IgM 实际上作为一种五聚体 IgMsJ(图 18－32)被分泌，在五聚体中 J 是主要连接的多肽(和 J 片段无关)以二硫键的形式和 μ 链连接。蛋白质的分泌是伴随着体液免疫反应进行的。

μ 重链的译本 μm(膜上的 μ 链)和 μs(分泌的 μ 链)仅在 C 末端有差异。在 μm 链末端有疏水序列，它可使 μ 链固定在膜上。在 μs 链中这一疏水区被一个短的亲水序列所取代，这种替代使 μ 链可以越过膜。C 端的改变是通过选择性的剪切完成的。剪切由 RNA 3′端控制(图 18－33)。

在膜结合阶段 RNA 在外显子 M2 后面终止，通过剪接由 6 个外显子构成恒定区。前 4 个编码区编码恒定区的 4 个功能区，后面 2 个 M1 和 M2 编码 41 残基组成的疏水 C－末端区和一个不翻译的拖尾序列。在外显子 4 中 5′端的剪接区和 M1 前端 3′剪接区连接。在分泌阶段，RNA 在外显子 4 的后面终止，在这个外显子中，5′剪接区不再起作用，这使外显子延伸了 20 个密码子。

在另一些恒定区也发现了与从膜蛋白到分泌型这种相似的转变。外显子结构的保持表明机制是相同的。Ig 的类型是由任何一个细胞合成的。与此不同的是 IgM 和 IgD 在成熟的 B 淋巴细胞中通过刺激而产生。这两种 Ig 除了重链的 μ 和 δ 恒定区替换以

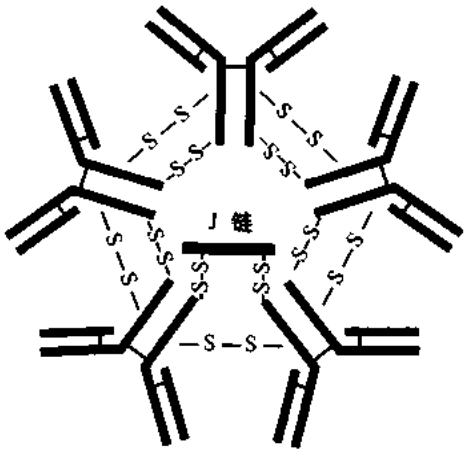


图 18－32　IgM 五聚体分子通过 J 多肽链与二硫键结合在一起。

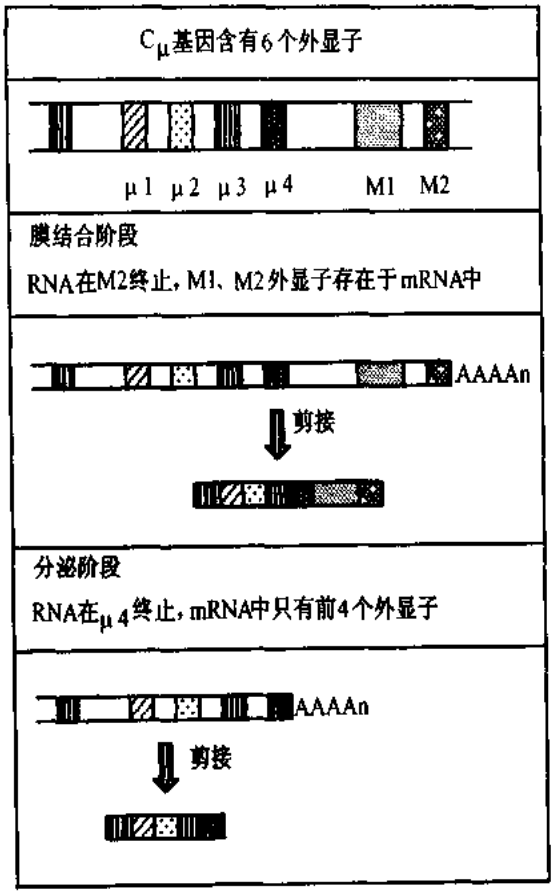


图 18－33　膜 IgM 和分泌 IgM 差异的来源。(仿 B.Lewin:Genes Ⅵ,1997,Fig.33－18)

外其他都是相同的。看来像是 RNA 选择性加工的结果。δ 恒定区紧靠 μ 区，有时转录延续穿过此区，*VDJ* 外显子可能被剪接成一系列 δ 外显子。

第三节　染色质水平的调控

一、异染色质化

在染色质中可分为**常染色质**(euchromatin)和**异染色质**(heterochromatin)，它们在细胞中凝聚的时期不同，异染色质是包装成 20～30nm，不具有转录活性的染色质。异染色质又分为**组成性异染色质**(constitutire heterochromatin)和**兼性异染色质**(facultatire heterochromatin)。前者是指在各种细胞中，在整个细胞周期内都处于凝聚状态的染色质，如着丝粒，端粒，核仁形成区等。后者指在某些特定的细胞中，或在一定的发育时期和生理条件下凝聚，由常染色质变成异染色质。这本身也是真核生物的一种表达调控的途径。

如水蜡虫(*Pseudococcus nipae*)(2n＝10)在体细胞里来自父本的 5 条染色体依次被异染色质化，在精子形成时丢失，只保留来自母本的 5 条染色体。

细胞质中存在着某些物质可以使染色体凝聚，使水蜡虫的父源染色体异染色质化。在哺乳动物中，细胞质某些调节物质也能使两条 X 染色体中的一条异染色质

化。只有一条 X 染色体具有活性，这样使得雌、雄动物之间虽 X 染色体的数量不同，但 X 染色体上基因产物的剂量是平衡的，这个过程就称为**剂量补偿**（dosage compensation）。

我们知道正常的男性是 XY，而正常的女性 XX。在正常女性的细胞核中有一团高度凝聚的染色质，而在正常男性的细胞核中都没有，它就是失活的 Xx 染色体，L. Barr（1949 年）首先发现它。因此它命名为**巴尔小体**（Barr body）（图 18－34）。在正常的 XX

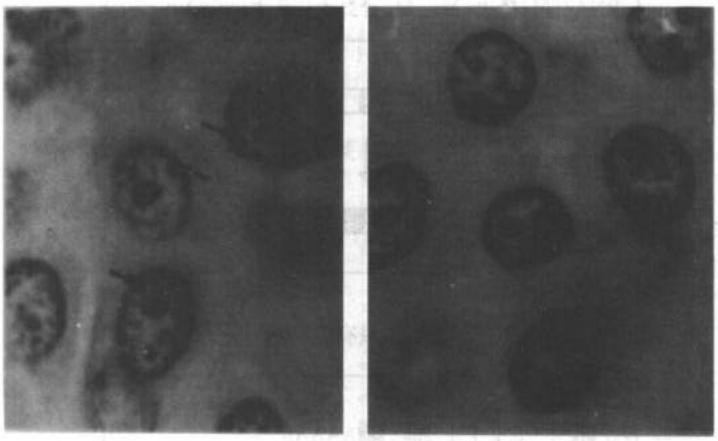


图 18－34　巴尔（Barr）小体：（a）正常女性细胞（XX）的核，箭头指向巴尔小体；（b）正常男性细胞（XY）的核，表明无巴尔小体。（转引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 3－16）

个体的体细胞中有一个巴尔小体。在正常 xy 个体中只有一 条 X 染色体和一条 y 染色体，而没有巴尔小体。在带有多条 X 染色体的个体中，只有一条 X 染色体是有活性的。巴尔小体的数目为 X 染色体的条数减 1（nx－1）。1961 年 Mary Lyon 提出了**莱昂假说**（Lyon hypothesiis）解释了巴尔小体的概念。她提出巴尔小体是一个失活（或大部分失活）的 X 染色体。她的假设解释了带有 X 染色体畸变的个体为什么能得以幸存。其主要论点如下：①巴尔小体是一个失活的 X 染色体，失活的过程就称为**莱昂化**（lyonization）；②在哺乳动物中，雌性个体细胞中的两个 X 染色体中有一个 X 染色体在受精后的第 16 天（受精卵增殖到 5000～6000，植入子宫壁时）失活；③两条 X 染色体中哪一条失活是随机的；④X 染色体失活后，在细胞继续分裂形成的克隆中，此条染色体都是失活的；⑤生殖细胞形成时失活的 X 染色体可得到恢复。

很多的实验证据都支持莱昂假设。如人类有一种 X－连锁的异常叫做无汗性外胚层发育不良（anhidrotic ectodermal dysplasia），本病主要表现为毛发稀少，牙齿发育异常，无汗或少汗，以及表皮和附件异常。杂合女性表现出来有齿和无齿颚区的嵌镶以及有汗腺和无汗腺皮肤的嵌镶。这两种嵌镶的位置在个体之间明显不同，这是由于发育期一条 X 染色体随机失活所致（图 18－35）。

三色猫（又叫做三玳瑁猫）也是一个很好的例子。雌性的三色猫腹部的毛是白色的，背部和头部的皮毛由桔黄色和黑色斑组成，十分漂亮。这种雌猫是一个 X－连锁基因杂合体，X－连锁的 *b* 基因控制橙色（orange）毛皮，其等位基因 *B* 是控制黑色的毛皮。带有 *b* 基因的 *X* 染体若失活，*B* 基因表达产生黑色毛斑，若带有 *B* 基因的 *X* 染色体若失活，b 基因表达则产生橙黄色毛斑。

X－连锁的 6－磷酸葡糖脱氢酶（G－6－PD）的测定也为莱昂假设提供了有力证据。女性虽然有两条 X 染色体，但其 G－6－PD 活性和男性的相同，表明其 X 染色体的总量有一半是失活的，这正好说明了剂量补偿作用。G－6－PD 有 A，B 两种类型，二者

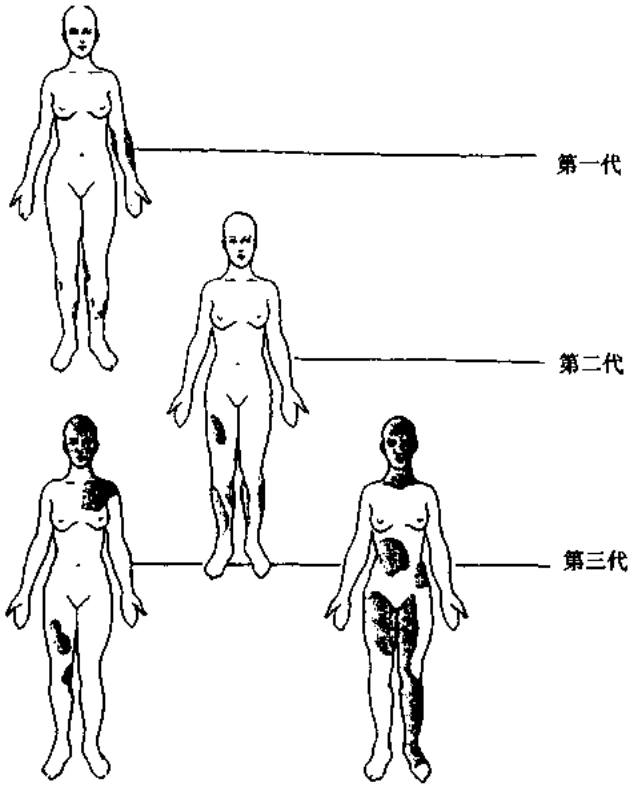


图 18－35　性连锁的无汗腺外胚层发育异常基因在 3 代女性中出现体细胞嵌镶现象。受累者缺乏汗腺。体表无汗腺区域用深色表示。图中显示各代女性都有无汗腺区，但分布各不相同。
（转引自 Griffiths A. J. F. *et al*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 3－30）

之间只有一个氨基酸的差异，但电泳带的迁移率不同，A 带比 B 带移动得快一点。它们分别由一对等位基因 Gd^A 和 Gd^B 编码的。Gd^A 或 Gd^B 纯合的女人从各种组织取样进行电泳分析都只出现一条电泳带。Gd^A/Gd^B 杂合的女人电泳的结果却会出 A、B 两条带。但用胰酶处理杂合体的皮肤细胞，使其分成单个细胞，然后进行克隆培养，每个克隆都来一个单细胞，再从各个克隆中取样进行电泳发现，每个克隆只出现一条电泳带非 A 即 B，非 B 即 A，而绝不出现二条带（图 18－36）。表明细胞中虽有一对等位基因 Gd^A/Gd^B 的存在，但由于有一条 X 染色体失活，使其上的等位基因不能表达，所以只出现一条带。虽然大量的证据表明了莱昂假设的正确性，但仍存在一些问题。

比如既然女人只有一条 X 染色体是有活性的，那么 XXX 和 XO 的女性也只有一条 X 染色体有活性，那么为什么会出现异常呢？1974 年 Lyon 又提出了新莱昂假说，认为 X 染色体的失活是部分片段的失活，这就能很好地解释以上的矛盾。现在已经查明在失活的 X 染色体上决定一种血型的 Xg 基因就仍然保持着活性。

二、组蛋白与非组蛋白

前面我们已讨论过真核的染色体是由 DNA 与组蛋白、非组蛋白构成的复合物。基因转录的调节最终必定涉及特殊 DNA 的核苷酸序列，这些序列可被相应的调节物分子所识别。

从进化的意义上说组蛋白是极端保守的，在各种真核生物中它们的氨基酸序列，结构和功能都十分相似。此外在生物体中 5 种不同类型的组蛋白都是以恒定的方式沿着 DNA 排列。显然组蛋白沿着 DNA 均匀分布所产生的系统不可能对成千上

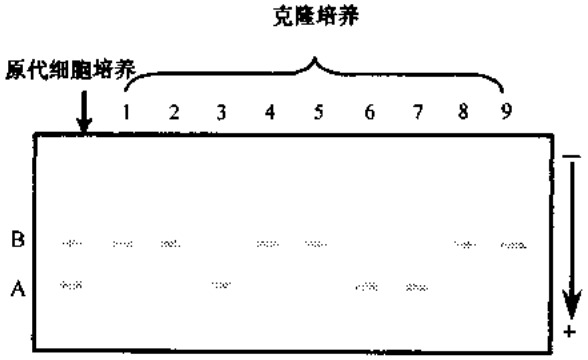


图 18－36　AB 表型女人培养细胞的 G－6－PD 电泳图。
（参考 Davidson *et al*.：Proc. Nat. Acad. Sci. 1963, 50: 481－485）

万个基因的表达进行特异控制。组蛋白可被修饰，如甲基化，乙酰基化和磷酸化。现在尚未发现组蛋白的修饰作用与染色质的功能之间有何明确的关系，但修饰可改变它们与 DNA 的接合能力。若被组蛋白覆盖的基因要表达。那么组蛋白必须要被修饰，使其和 DNA 的结合由紧变松，这样 DNA 链才能和 RNA 聚合酶或调节蛋白相互作用。因此组蛋白的作用本质上是真核基因调节的负控制因子，即它们是基因的抑制物。

如果组蛋白不是负责打开特异基因的分子，那么这个作用一定是由非组蛋白来控制。非组蛋白不仅数量多（约有数以千计的不同类型），其在不同细胞中数量和分布都是十分有意义的。因此人们预期具有不同功能的细胞中表达的基因排列是不同的。不同细胞间非组蛋白的变异有力地表明非组蛋白在基因表达的调节、细胞分化的控制以及生物的发育中将起着很重要的作用。

非组蛋白特异调节作用的的证据来自染色质重建实验。来自一个细胞的染色体可以解离成 DNA，组蛋白和非组蛋白三种组分。来自不同类型染色质组份能够混合和匹配也可以重新建成染色质。若将无转录活性细胞中的 DNA、组蛋白与具有转录活性细胞中的非组蛋白相混合，将会产生转录活性（图 18－37）。

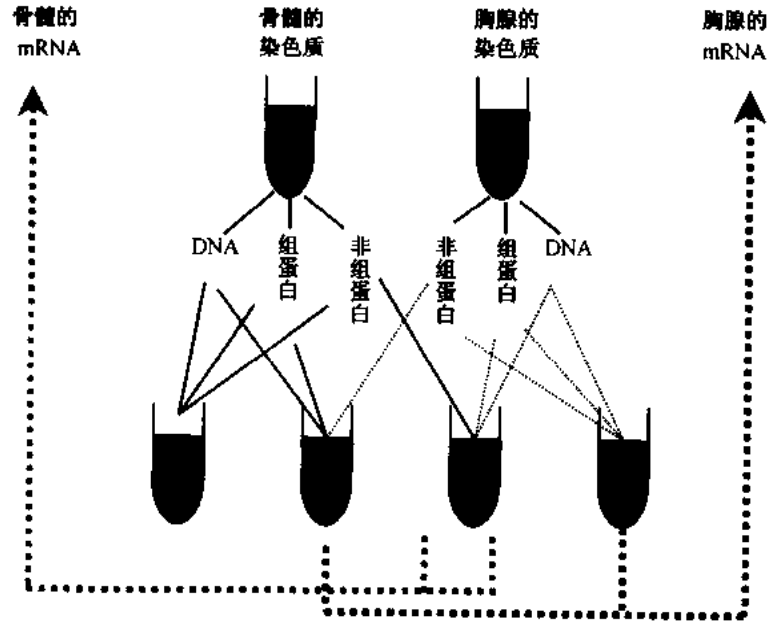


图 18－37　染色质重建实验。

证据表明有转录活性的染色质要比无转录活性的染色质外形要松散得多。这是核小体中 DNA 的高度紧密的二级结构和三级结构打开了，使 DNA 易于为调节蛋白和 RNA 聚合酶所接近。经过几年的大量研究提示了活性基因与无活性基因之间的重要差别，不仅涉及到包裹在 DNA 外面的蛋白质，而且还涉及到 DNA 蛋白结构的折叠和超螺旋。

三、DNase 的敏感性和基因表达

当一个基因成为转录活性状态时，含有这个基因的染色质区域对 DNase Ⅰ（一种内切酶）降解的敏感性要比无转录活性区域高得多。这是由于此区域染色质的 DNA 蛋白质结构变得松散，DNase Ⅰ易于接触到 DNA 之故。

一个很好的研究系统是鸡的珠蛋白和卵清蛋白系统。来自鸡胚红细胞的核中的珠蛋白的基因对 DNase Ⅰ的降解是敏感的，而编码卵清蛋白的基因在红细胞中是不表达的，它对 DNase Ⅰ的降解也是具有抗性的。相反，在母鸡输卵管细胞的核中，卵清蛋白的基因是具有转录活性的，但却不能产生珠蛋白的 RNA，卵清蛋白基因的序列对 DNase Ⅰ 的降解要比珠蛋白基因的序列敏感得多。

此实验以克隆的特异基因为探针，应用 Southern 杂交技术来论证 DNA 序列对 DNase Ⅰ敏感性的改变，图 18－38 表示用这种方法对鸡的珠蛋白基因和卵清蛋白基因的检测。用不同浓度的 DNase Ⅰ来处理样品，然后从各染色质样品中抽取 DNA，再用限制性酶 *Bam*H Ⅰ来降解，它可在珠蛋白基因两侧中的任一侧剪切，两个 *Bam*H Ⅰ位点在染色体 DNA 上相隔 4.5kb，此 DNA 片段经琼脂糖电泳并通过 Southern 吸印技术转移到尼龙膜上，此膜和放射性珠蛋白探针进行分子杂交，来确定珠蛋白基因 DNA 片段的大小，从而可以推断是否被 DNAase 降解过。

实验结果如下：红细胞中提取的染色质，珠蛋白基因含有的 DNA 片段看来随着 DNase Ⅰ浓度的增加 4.5kb *Bam*H Ⅰ－*Bam*H Ⅰ逐渐消失，表明含有珠蛋白基因的染色质片段对 DNase Ⅰ的降解是敏感的，即珠蛋白的基因区域染色质组成是松散的，珠蛋白基因 DNA 序列很容易和 DNase Ⅰ接触。从对照组细胞（不产生珠蛋白）中分离的染色质其 4.5kb *Bam*H Ⅰ DNA 片段不被任一浓度的 DNase Ⅰ所降解。这表明不产生珠蛋白的细胞中相关的染色质是高度凝聚的，因而珠蛋白的基因序列不易和 DNase Ⅰ接触，从而不发生降解。用克隆的卵清蛋白基因为探针，来分析红细胞中的卵清蛋白基因（用 *Bam*HI 酶来处理，可切成 20kb DNA 片段）表明卵清蛋白基因对 DNase Ⅰ的降解不敏感，20－*kb Bam*H Ⅰ－BamH Ⅰ DNA 片段在各种浓度的 DNase Ⅰ作用下依然存在（图 18－38），表明基因组中这一区域并不变得松散。

很多这类实验得出的结论几乎都是有转录活性的基因对 DNase Ⅰ的敏感性增加，DNase Ⅰ敏感区的范围随着基因序列的不同而变化，从基因周围几个 kb 到两侧 20kb 大小不等。

仔细分析具有转录活性基因周围的 DNA 区域，表明有一个中心区域存在，称为**超敏感区域**（hypersensitive region）或**超敏感位点**（hypersensitive site），它对 DNase Ⅰ是高敏感的。这些位点或区域将首先受到 DNase Ⅰ的剪切。由于对 DNase Ⅰ的敏感性反映了染色质中 DNA 的有效性，我们将这些位点描述为染色质中特殊 DNA 暴露区域，一般在转录起始点附近，即 5′端启动子区域，少部分位于其他部位如转录单元的下游。

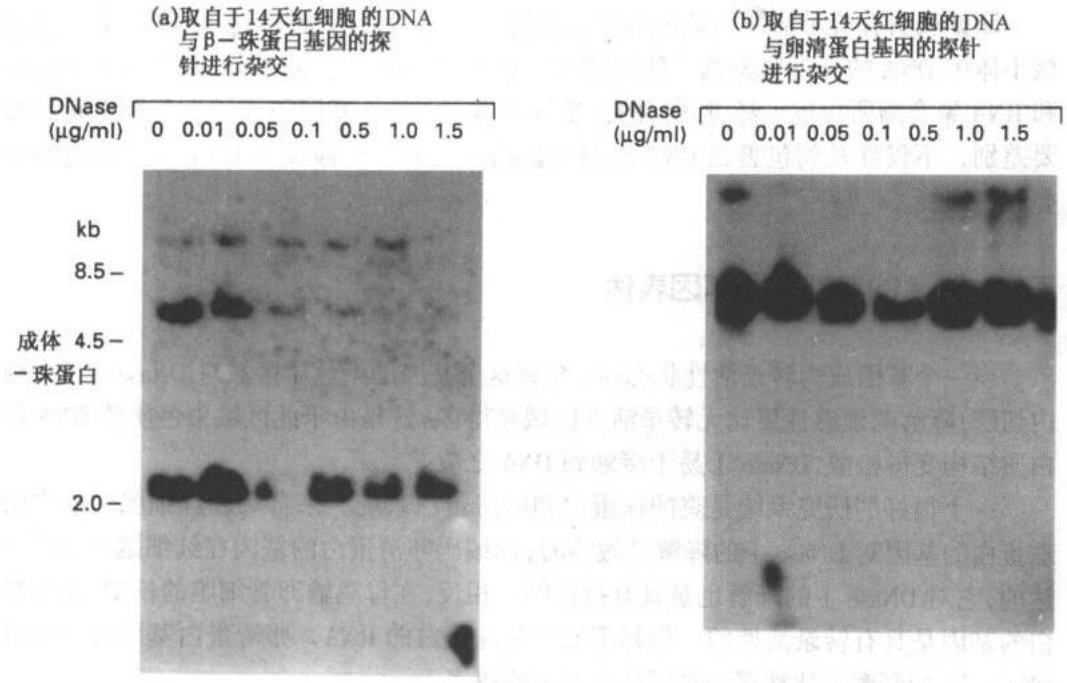


图 18－38　基因的表达和对 DNase 敏感性之间的关系。(a)在红细胞中，珠蛋白 DNA 对 DNase Ⅰ敏感；(b)在红细胞中，卵清蛋白 DNA 对 DNase Ⅰ不敏感。(转引自 B. Lewin: Genes Ⅵ, 1997, Fig27－42)

超敏感位点建立于转录起始之前，转录的诱导物除去后虽然超敏感位点依然存在，但转录却很快停止，说明超敏感位点的存在是转录起始的必要条件，但不是充分条件。

超敏感位点是一段长 200bp 左右的 DNA 序列，这些区域是低甲基化区；并可能有局部解链的存在；不存在核小体结构或结构不同寻常，此区因 DNA 的裸露易和多种酶或特异的蛋白质结合，也就是说易于和反式作用因子结合。

染色质对 DNase Ⅰ的敏感性与 2 种非组蛋白有关，那就是高泳动蛋白 14 (**HMG14**, high-mobility group)和 **HMG17**，这是两种高丰富度小分子量(30kDa)的蛋白，活化染色质中每 10 个核小体就结合一个 HMG 分子。当从红细胞中提出这两种蛋白时，珠蛋白基因不再对 DNase Ⅰ出现敏感，当将这两种蛋白重新加入到此系统中，敏感性又可得到恢复。

HMG 蛋白的 C 端含活性氨基酸，可与核小体核心组蛋白的碱性区域相结合。HMG 蛋白的 N 端 1/3 的区域氨基酸序列与 H1 和 H5 的 C 端十分相似。而 HMG 在核小体上位置与 H1 相近，HMG 和组蛋白相互作用。可以竞争性取代 H1 或 H5，核小体缺乏 H1 和 H5 将使染色质变得松散，成为具有转录活性的状态。

除 HMG 以外可能还有别的因素会影响到染色质的活性，当把上述两种 HMG 加入鸡脑细胞的染色质中，其中珠蛋白基因所在区域并不表现出对 DNase Ⅰ的敏感，表明红细胞和脑细胞还存在其他因子的差异，这些因子在盐抽提后仍留在染色质中，它们决定了红细胞珠蛋白基因区域在 HMG14 和 HMG17 存在时对 DNase Ⅰ敏感。

四、座位控制区

在人类中 α－珠蛋白和 β－珠蛋白的相关基因分别以基因簇的形式存在，基因簇

中的各个基因在胚胎发育和成体发育不同时期进行表达。这些基因有大量的调节元件，人们已对这些元件进行了较为深入的研究。在成人的β－珠蛋白基因中，调节序列位于基因的5′和3′端；在启动子区域包含有正、负调控元件；在基因的下游还有附加的正调控元件。含有这些控制区域的人类β－珠蛋白基因在转基因小鼠中不表达，看来还需要另一些调节序列。通过DNase Ⅰ超敏感位点在这个基因簇的上游和下游已鉴别出具有附加调节功能的区域。图18－39表示ε基因上游20kb的范围内含有一组5个超敏感位点，在β基因下游30Kb外有一个超敏感位点。将各种结构转染到小鼠的红细胞中表明，5′端的各位点之间的序列被切除后没有大的影响，但若切除任何一超敏感位点就降低整个的表达水平。小鼠的β－珠蛋白基因簇含有一个与人类β－珠蛋白的基因族上游相似的区域，且显示出很好的保守性。

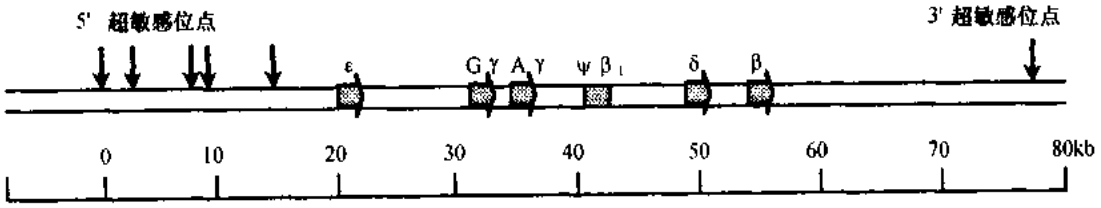


图18－39 在珠蛋白功能区的两端被超敏感位点所标记。5′端一组位点含有LCR，此对于这个基因簇中的各基因的功能来说都是必要的。（仿B.Lewin:Genes Ⅵ，1997，Fig.29－18）

5′端的调节位点是初级调节物，这一簇超敏感位点被称为**座位控制区**（LCR，locus control region）。3′位点的出现取决于5′位点，5′LCR的缺失将抑制3′位点的出现。此LCR对珠蛋白基因簇中每个基因的表达是绝对需要的。每个基因的进一步调节是由它们自己特有控制机制来进行的。这些控制有的是"自治的"：ε和γ基因的表达本质是这些位点和LCR的连接，另一些控制是依赖在基因簇中的位置，这就意味着基因簇中的基因排列序列对调节来说是重要的。

含有珠蛋白基因的整个区域及延伸区将超过这一区域的本身，构成一个染色体功能区，此功能区通过对DNase Ⅰ敏感性的增加而加以确定。5′LCR缺失可恢复整个区域对DNase Ⅰ的正常抗性，可能LCR对整个区域的"打开"并使其有效转录是必须的。此功能区从一个核基质结合区MAR（matrix association region）5′延伸到相邻基因那一边的超敏感位点，看来离β基因约100kb，此MAR限定在功能域的5′端，3′端由什么来限定尚不清楚。MAR多半在基因簇附近。

通过分析黑腹果蝇基因获得了有关功能区的一些信息（图18－40），位于18kb区域中的两个基因编码热激蛋白（hsp，heat-shock protein），此区位于87A7带。此区的两端是**有特化染色质结构**（scs和scs′：specialized chromatin structures）。由350bp构成的*scs*或200bp构成的*scs′*对DNase Ⅰ有高度的抗性，而在其两侧又具有间隔的100bp超敏感位点，在热休克条件下剪切模式会发生变化。

*scs*元件使*hsp70*基因与周围区域的因子隔离开，若我们将scs单位取出放到白眼基因（white－eye gene）的两侧，那么这个基因无论放到基因组的任何位置都能行使功能，甚至放在异染区。

看来scs单位在控制基因表达中并不能起到正、负调节的作用，而是限制一些因子从一个区域到另一个区域。若附近区域有抑制因子，scs元件可以阻断这个因子作用的延伸，因此对基因表达来说是必要的。当scs元件缺失时，此带区的功能会消失。看来它们的作用是作为一种**"隔离子"**或**"绝缘体"**（insulator）。

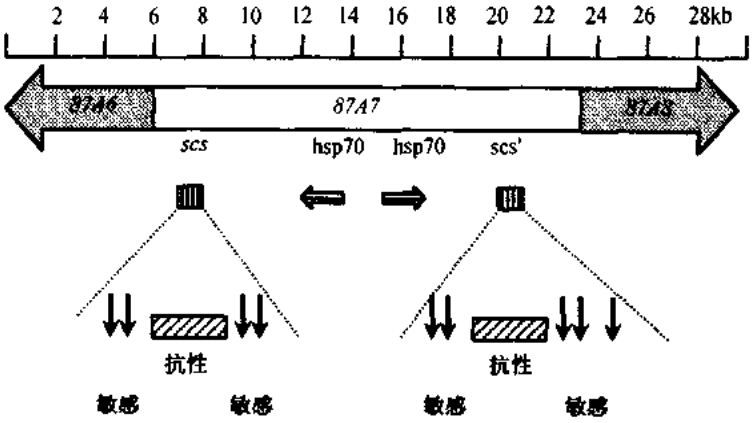


图 18－40　在果蝇的基因组中特化的染色质结构。在功能区的末端产生超敏感位点，屏蔽了基因，使其不受两侧序列的影响。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 2－19）

五、核基质蛋白

染色质并不漂浮在核内，而是结合在**核基质**（nuclear matrix）上，核基质是一种 3～5nm 的网络纤维，又称为**骨架蛋白**（scaffolding protein）。这种结合是特异性的，如卵白基因与小鸡卵巢的细胞核基质结合，而不与小鸡肝脏或红细胞的核基质结合；珠蛋白基因不与卵巢核基质结合，而与红细胞的核基质结合，这种特异性的结合对于控制基因的活性是有用的。

为什么会出现特异性的结合呢？这是由于不同的基因存在着不同的**核基质结合区**（matrix association region，**MAR**）。启动子与增强子相距甚远，这种远距离效应是通过形成**侧环**（looping out）来实现的。侧环的附着基部（primary loop anchorage）有种特殊的序列就是 MAR，或称为**骨架附着区**（scaffold attached region，**SAR**）。如卵清蛋白的基因具有特殊的 MAR，可以和卵巢核基质结合形成侧环，使增强子和启动子彼此靠近，通过一系列转录激活因子可使基因表达，但这种 MAR 却不能和红细胞的核基质结合形成适当的侧环，也就无法表达。

免疫球蛋白的重链基因 IgH（γ2b）有一段内含子由 992bp 的 *Xba* 片段组成（图 18－41），含有两个负调节序列 IN1 和 IN2，它们在成纤维细胞中无活性，但在骨髓瘤细胞中却有活性。中心部分是增强子区，是 B 细胞特异的正调节区。在非淋巴细胞中 IN1 和 IN2 和反式因子结合，且都在 MAR 区，由于蛋白质的相互作用形成侧环，使起正调节作用的增强子区夹在环的中间，无法靠近启动子，故不能发挥其“增强”的功能。

如果我们把在不同系统中发现的各种结构放在一起，我们就可以较全面地了解染色体功能域所存在的一些特点，图 18－42 表示已介绍的一些功能域的结构。一个功能域（在果蝇的染色体上作为一条斑带）可能含有一系列的转录单位。此功能单位可能还需要一系列顺式作用结构。这儿有一种与 hsp70 带的 scs 相当的单位，我们可称之为“**限定子**”（delimiter）或“**绝缘子**”（insulator），可以停止激活因子或抑制因子传向这个功能单位。这可能和核基质附着位点相同，或与 MAR 分开，也可能和 LCR 相同或分离开的一种重要的结构，此结构的远距离功能对此单位中各个基因的表达都是重要的。这些元件已在功能区的一端或两端被发现了。

由于 α－珠蛋白基因簇有不同的组成，在这种组成中顺式－作用控制元件分布于广阔的区域中。α－基因的有些控制元件是位于 α－基因簇的外侧，有的在持续表达的

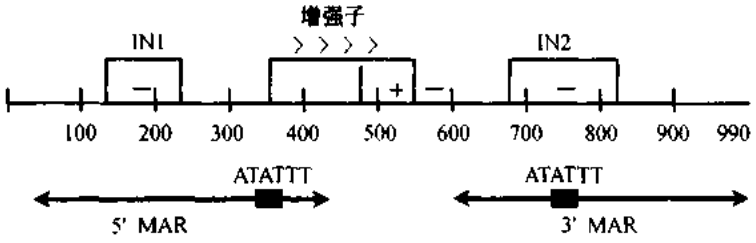


图 18-41　免疫球蛋白重链基因 IgH(γ2b)*X6a* Ⅰ的 MAR。

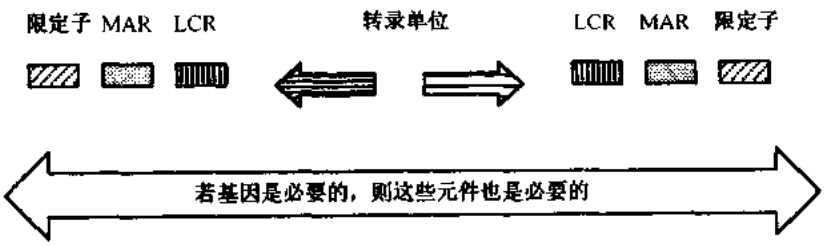


图 18-42　功能区可能有三种位点：限定子(阻止相邻功能区之间的影响)；MAR(将功能区附着在核基质上)；LCR 为起始转录所需。(仿 B. Lewin: Genes Ⅵ, 1997)

基因的附近，这样就不大可能限定各个与 α-珠蛋白相关的功能区孤立的表达。

第四节　DNA 水平的调控

一、DNA 的甲基化与去甲基化

DNA 上存在着细胞行使各种功能所需要的各种信息，在行使不同功能时又需在DNA上利用或消除有关的信息，这种利用和消除可能是由 DNA 可逆性甲基化提供的。动物细胞中 DNA CpG 位点发生 C 的甲基化，是一种复制后的加工形式，不同的时间空间甲基化的形式千变万化。总的来说有半甲基化：$\frac{5'^{m}CpG3'}{3'GpC5'}$，和甲基化：$\frac{5'^{m}CpG3'}{3'GpC^{m}5'}$。我们采用同裂酶(isoschizomers)来处理待测序列，比较相应切点就能弄清哪些位点甲基化了，哪些位点没有甲基化。例如 *Hpa* Ⅱ识别并切割未甲基化的 CCGG($C^{\downarrow}CGG$)，而 *Msp* Ⅰ识别无论是否甲基化的 CCGG($C^{\downarrow}CGG$ 或 $C^{\downarrow}C^{*}GG$)，切割的片段分别进行电泳(图 18-43)。*Hpa* Ⅱ只能切非甲基化的 CCGG，所以只有一个切点，产生两条带；而 *Msp* Ⅰ可以切甲基化和未甲基化的 CCGG，因此有两个切点，产生 3 条电泳带，比较带的长度就不难确定甲基化位点的位置了。

在一些不表达的基因中，启动区的甲基化程度很高，而处于活化状态的基因则甲基化程度较低。现以卵黄原蛋白Ⅱ基因为例来说明这个问题。卵黄蛋白并非在卵母细胞或受精卵中合成的，而是在成熟的雌性动物肝中先合成卵黄蛋白原(vetellogenin)，然后分泌到血液中，再运送到卵巢。发育中的卵(卵母细胞)选择性地将卵巢中的卵黄蛋白原加以吸取，再切割，装配成卵黄蛋白。雄性动物也有卵黄蛋白原基因，但不能表

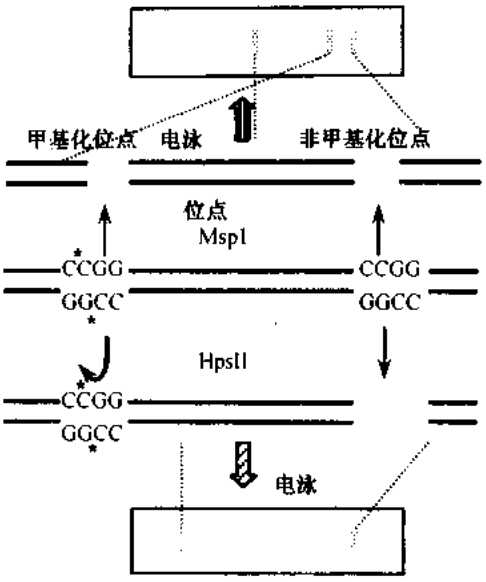


图 18－43　用同裂酶检测甲基化位点。

达。这是因为缺乏雌激素之故。其分子机制仍和甲基化有关。鸡的卵黄蛋白原基因的 5′调节区有 4 个 CpG 位点（图 18－44），C，D 位点是雌二酮受体蛋白复合物的结合部位。在雌激素处理后，在－612（C）的 *Hpa* Ⅱ位点处于低甲基化。在红细胞 DNA 的上面一股链上 4 个位点全部甲基化，下股链只有 50%甲基化，但经雌二酮处理后数小时内甲基化的形式发生急剧变化，上股链中 4 个位点全部去甲基化，与卵黄蛋白原 mRNA 相吻合，而另一股键上 4 个位点滞后 24 小时才去甲基化。在肝细胞中卵黄蛋白原基因中的 5′端调节区有部分半甲基化位点，而雌激素的处理只引起其中一条链迅速地去甲基化，另一条链暂时仍保持半甲基化状态，此基因在雄体中处于甲基化状态，没有活性。

酶对 CpG 位点的甲基化是有选择的，例如小鼠 β－珠蛋白基因 5′侧翼有 5 个 CpG

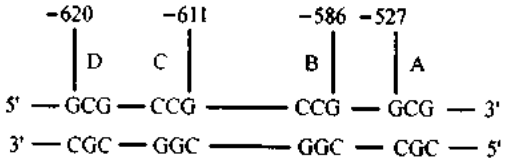


图 18－44　鸡卵黄蛋白原基因 5′端调节区有 4 个 CpG 位点。

位点，其中 4 个较集中。用纯化的 DNA 甲基转移酶在体外处理这段序列，结果各位点的甲基化程度各不相同，有的不发生甲基化，有的甲基化不明显，有的稍有甲基化，有的大量甲基化，无论是诱导的还是未诱导的细胞；无论是纯化酶，还是粗制酶都一样。表明还有其他的选择因素存在。

甲基化可使基因失活，去甲基化又可使基因恢复活性。5－氮胞苷（图 18－45）没有游离的 5′－H，因而不能接受甲基，若将它掺入 DNA 取代胞苷是可以改变基因的甲基化状态，而达到去甲基化。

Mohandas 等将含有一条 X 染色体易位（X/11 易位）的女性成纤维细胞和小鼠 A9 细胞（*HPRT*⁻）融合后用 HAT（培养基中加入了次黄嘌呤（H），氨基喋呤（A）和胸苷（T）），选择带有 *HPRT*⁺ 基因的细胞，而人类 X 染色体的长臂上带有此基因，若 X 染色体失活，这个基因也没有活性。选择出的融合细胞中一部分带有 2 条人的 X 染色体，一条有活性，一条是失活的。随着培养的延续，杂种细胞将继续排斥人类的染色体，在这种情况下再用硫基鸟嘌呤进行反选

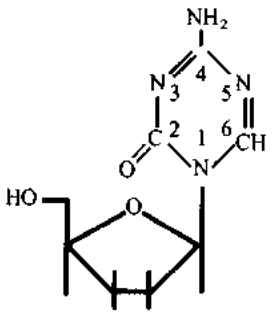


图 18－45　5′氮胞苷。

择(back-selected)，硫代鸟嘌呤是有毒性的，*HGPRT*⁺的细胞可使其掺入DNA而导致死亡，而*HGPRT*⁻的细胞反而可以得以生存，选择出来的细胞只有一条正常的X染色体(如未易位)，表明这条染色体是失活的，所以虽有*HGPRT*基因，但不表达。接着他们用5-氮胞苷处理细胞，然后再在HAT培养基中克隆，结果细胞增加了100倍。图18-46表明原来失活的X染色体经5-氮胞苷处理后恢复了活性，*HGPRT*基因变成活性状态，这样才能在HAT培养基中分裂。这个实验说明*HGPRT*基因在失活染色体上重新被活化是由于5-氮胞苷起到去甲基化的作用，也说明莱昂化不是抑制因子的持续存在，而是一条染色体被选择性甲基化之故。

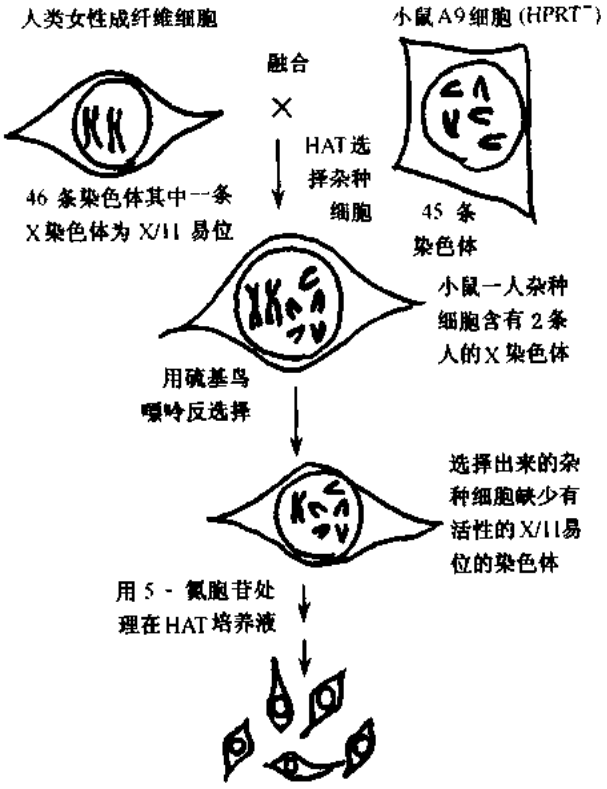


图 18-46　含有一条失活X染色体细胞的分离及5-氮胞苷使HPRT基因恢复活性实验。

第五节　转录水平的调控

一、顺式作用元件和反式作用因子

真核生物转录的调控也是调控的最重要的途经，其启动子存在着很多顺式元件(见12章)可分为以下四种类型：① 核心启动子成分，如TATA框；②上游启动子成分，如CAAT框，GC框，八聚体(octamet)，ATF结合位点；③远上游序列：如增强子，酵母中的UAS(upstream activator sequences)减弱子、静息子等；④特殊细胞中的启动子成分：如淋巴细胞中的Oct(octamer)和κB。这些元件都为不同的转录因子及蛋白质识别和结合(表18-6)来调节转录。

表 18 - 6　哺乳动物 RNA Pol Ⅱ 启动子上游转录因子结合的序列元件

组件	保守顺序	DNA 长度	结合因子	大小(Da)	丰度(/细胞)	分布
TATA box	TATAAAA	~ 10bp	TBP	27 000	?	普遍
CAAT box	GGCCAATCT	~ 22bp	CTF/NF1	60 000	300 000	普遍
GC box	GGGCGG	~ 20bp	SP1	165 000	60 000	普遍
Octamer	ATTTGCAT	~ 20bp	Oct - 1	76 000	?	普遍
Octamer	ATTTGCAT	23bp	Oct - 2	52 000	?	淋巴细胞
KB	GGGACTTTCC	~ 10bp	NF_KB	44 000	?	淋巴细胞
KB	GGGACTTTCC	~ 10bp	H2·TH1	?	?	普遍
ATF	GTGACGT	~ 20bp	ATF	?	?	普遍

B. Lewin: Genes Ⅴ., 1994, Table 29.3

在真核细胞中鉴别出大量的转录因子，有的结合在增强子区，如甾类 - 受体复合物；另一些结合在启动子的顺式作用元件上，与 RNA pol Ⅱ 结合成前起始复合体，使 RNA pol Ⅱ 起始转录。

根据靶位点的特点反式作用因子可以分为 3 类；①**通用反式作用因子**，在一般细胞中普遍存在，主要识别一些启动子的核心启动成分 TATA 框，如 TBP；识别上游启动子成分 CAAT 框，如 CTF/NF - 1；识别 GC 框如 SP1；识别八聚体核苷酸的 Oct - 1 等；②特殊组织与细胞中的反式作用因子，如淋巴细胞中的 Oct - 2；③和**反应性元件**(response elenents)相结合的反式作用因子。反应性元件如 **HSE**(热休克反应元件，heat shock response element)，**GRE**(糖皮质激素反应元件 glucocorticoid response element)；**MRE**(金属反应元件，metal response element)；**TRE**(肿瘤诱导剂反应元件，tumorgenic agent response element)；**SRE**(血清反应元件 serum response element)(表 18 - 7)。

表 18 - 7　结合于反应元件的特殊转录因子

调节剂	组件	保守顺序	DNA 长度	因子	大小(Da)
热休克	HSE	CNNGAANNTCCNNG	27bp	HSTF	93000
糖皮质激素	GRE	TGGTACAAATGTTCT	20bp	受体	94000
弗波酯	TRE	TGACTCA	22bp	AP1	39000
血清	SRE	CCATATTAGG	20bp	SRF	52000

B. Lewin: Genes Ⅴ., 1994, Table 30.1

反应元件是启动子或增强子的上游元件，它们含有短的保守序列。在不同的基因中反应元件的序列密切相关，但并不一定相同，离起始点的距离并不固定，一般位于上游小于 200bp 处，有的也可以位于启动子或增强子中。

反式作用因子通过以下不同的途径发挥调控作用：蛋白质和 DNA 相互作用；蛋白质和配基结合；蛋白质之间的相互作用以及蛋白质的修饰。

1. 蛋白质直接和 DNA 结合

(1)螺旋 - 转角 - 螺旋(helix-turn-helix)

在很多细菌的调节蛋白中有这种相同的模体存在，如 λ 的 C Ⅰ 蛋白，Cro 蛋白

(见17章),在酵母中涉及交配型的 a1 和 a2 蛋白也是以这种形式和 DNA 结合的。在高等真核生物中的 Oct-1 和 Oct-2 也属于此类型;在果蝇早期胚胎发育中决定体节的一些基因如 *bicoid* 编码的蛋白称为**同源异形蛋白** (homeotic proteins) 也含有相似的 DNA 结合功能域。此序列称为**同源异形盒**(homeotic box)和原核的 HTH 有一些差别。螺旋转角旋有 3 个螺旋,螺旋 3 是识别和结合 DNA,一般结合于大沟;螺旋 1 和 2 和其他蛋白质结合(图 18-47)。

(2)锌指结构

锌指 (zinc fimger) 这个名称来源于它的结构,它由一小组保守的氨基酸和锌离子结合,在蛋白质中形成了相对独立的功能域,像一根根手指伸向 DNA 的大沟。两种类型的 DNA 结合蛋白具有这种结构,一类是锌指蛋白,另一类是甾类受体。典型的锌指蛋白有一组锌指(图 18-47b),单个的锌指的保守序列是: Cys - X_{2-4} - Cys - X_3 - Phe - X_5 - Leu - X_2 - His - X_2 - His, 根据锌结合位点的氨基酸又把锌指分成为 2Cys/ 2His 和 2Cys/ 2Cys 两类, 前者为Ⅰ型, 后者为Ⅱ型锌指, Ⅰ型结构 "指" 的本身由 23aa 组成, "指"与"指"之间通常由 7-8aa 连接。

转录因子 TF Ⅲ A(RNA pol Ⅲ转录 5S RNA gene 时所需) 的锌指结构组有 9 个 "指"组成串联重复。在转录因子 SPI 中只有 3 个"指";它和用 RNA pol Ⅱ的很多启动

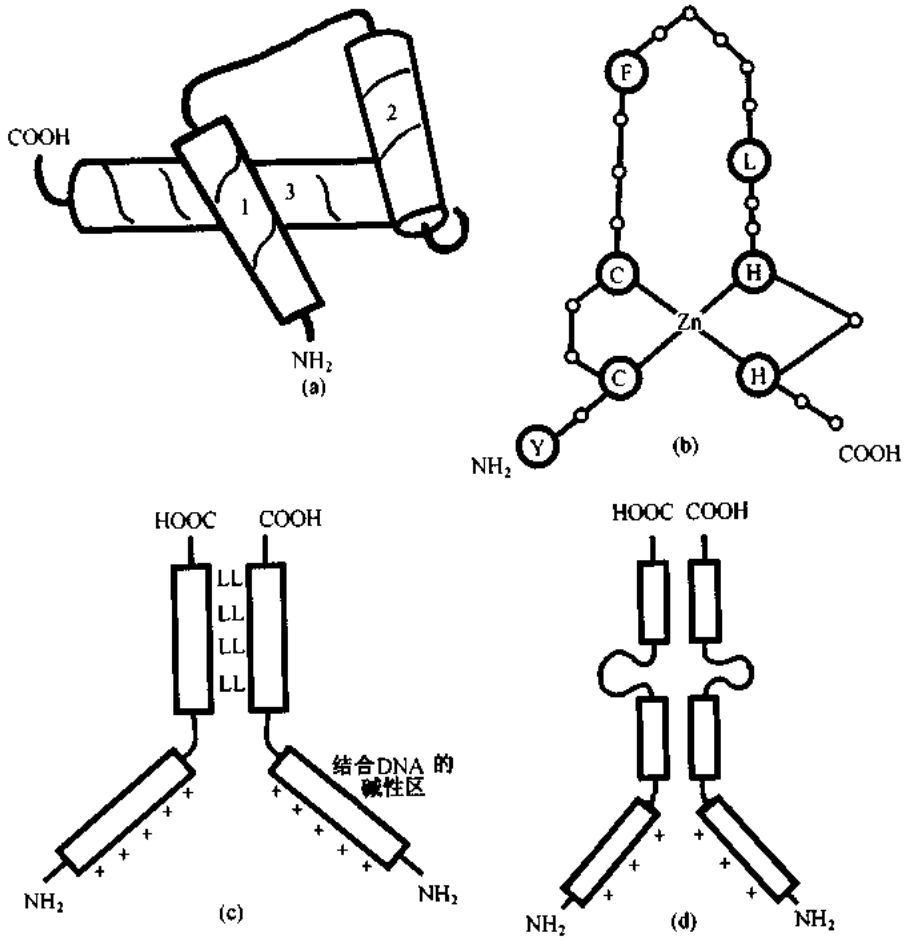


图 18-47 DNA 结合功能域的模型。(a)螺旋转角螺旋(b)锌指(C_6H_2);(c)亮氨酸拉链;(d)螺旋-环-螺旋。
(摘自 J.D.Watson *et al*.: Recombinant DNA, 2nd ed., Copyright 1992 by Scientific American Books)

子的起始有关。还有其他的一些 DNA 结合蛋白都属于此类锌指蛋白(表 18－8)。2Cys 在反向平行的 β－折叠中,而 2His 在 α 螺旋中(图 18－48),I 型锌指属于较原始的类型。

Ⅱ型锌指的保守序列为: Cys－X_2－Cys－X_{4-3}－Cy_3－X_2－Cys。带有此类锌指结构的调节蛋白如表 18－9 所示。这类蛋白常有非重复的指,和 I 型锌指结构不同。此类锌指的 DNA 的结合位点是短的回文序列。突变分析表明调节蛋白和 DNA 结合的 区域包括锌指模体。两型锌指是有差别的。用甾类受体做实验，表明如经突变将后两个 Cys 变成 His 的话，这种受体就不能激活靶基因。因此这两类锌指是不能互相转变的。

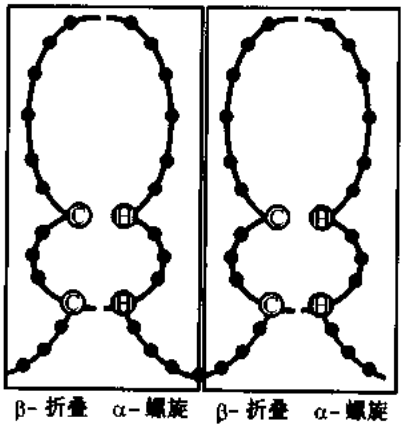


图 18－48　锌指可以形成 α 螺旋插入大沟,与其相连结的 β－折叠在另一侧。(仿 B. Lewin : Genes Ⅵ, 1997, Fig. 29－4)

糖皮质激素和雌激素受体都有两个“指”,每个指在 Cys 四分体的中央带有一个锌原子。两个指形成 α－螺旋，彼此折叠成一个大的球状功能域。α－螺旋的芳香族和与螺旋相连的 β 折叠一道形成一个疏水中心。N－端螺旋的一侧和 DNA 的大沟接触 (图 18－49)，两个糖皮质激素受体形成二聚体和 DNA 结合,每个受体负责和连续的大沟接触。每个“指”控制着受体的重要特点。图 18－49 表明在第一个“指”的右侧控制 DNA 的结合,在第二指的左侧控制着形成二聚体的能力。

表 18－8　带有 Cis/His 锌指结构的转录因子和 DNA 结合蛋白

蛋白	来源	大小	指	结合的 DNA	靶	启动子	功能
TFⅢA	哺乳动物	37 000	9	50bp	5S 基因	pol Ⅲ	5S 基因转录因子
SP1	哺乳动物	105 000	3	10bp	GC box	pol Ⅱ	一般的转录因子
ADR1	果蝇	150 000	2	22bp	ADH_2 基因	pol Ⅱ	激活 ADH 基因
Kruppel	果蝇	60 000	5	?	?	?	胚的分节
驼背基因产物	果蝇	80 000	4＋2	?	?	?	胚的分节

参考 B. Lewin: Genes Ⅳ, 1990, Table 29.8

表 18－9　Cys2/Cys2 型的调节蛋白

蛋白	大小	指	靶
糖皮质激素受体	94 000	2	GRE 中 20bp
雌激素受体	66 000	2	ERE 中 20bp
GAL4(酵母)	99 000	1	UAS 中 17bp
腺病毒 E1A	~30 000	1	?

参考 B. Lewin: Genes Ⅳ, 1990, Table 29.9

直接的证据表明第一锌指的 DNA 结合区是通过“特异的交换”实验获得的，即雌激素受体的锌指被缺失,而以糖皮质激素受体取而代之。这个新的蛋白可识别 GRE 序列(平常是糖皮质激素的靶序列),而不识别 ERE 序列(雌激素的靶序列)。GRE 和 ERE 的序列有些相似,进一步交换实验表明二者的第一个锌指基部的 2 个氨基酸残基有差别。

(3)亮氨酸拉链(Leucine zipper)

亮氨酸拉链是一种富含亮氨酸的蛋白链形成的二聚体模体。它本身形成二聚体同时还可以识别特殊的 DNA 序列。

一个亲水的 α - 螺旋在其表面的一侧有疏水基团(包括亮氨酸),另一侧表面带有电荷。亮氨酸拉链蛋白通过疏水界面上的亮氨酸形成二聚体。在肽链上两 Leu 之间相隔 7 个氨基酸,这样使它们在 α - 螺旋中排列成一条直线(图 18 - 47C)。亮氨酸拉链的侧翼是 DNA 结合功能区,含有很多的 Lys 和 Arg。二聚体的形成对于 DNA 的结合是必要的。很多原癌基因,例如 c-*jun* 和 c-*fos* 都编码亮氨酸拉链蛋白。它们本身相互作用,形成**同二聚体**(homodimers)或**异二聚体**(heterodimers)。c - Jun 和 c - Fos 蛋白的异二聚体的结合比它们同二聚体更为坚固。Myc, C/ EBP(结合 CAAT box 和 SV40 核心增强子的一种因子),AP1 蛋白(异二聚体,结合于 SV40 的增强子)等都具有亮氨酸拉链结构。过去认为二聚体之间亮氨酸与亮氨酸是交错连接,像"拉链"一样。而现在认为亮氨酸与亮氨酸是相对连接的。

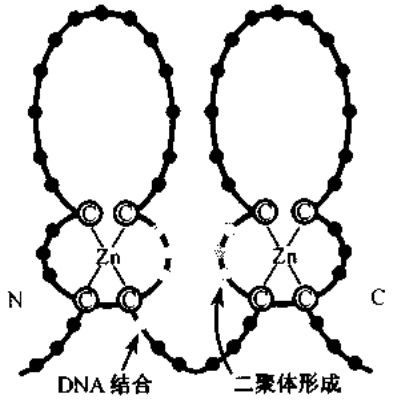


图 18 - 49 甾类受体的第一个"指"含有 DNA 结合的特异位点。第二个指含有特异的二聚体形成位点。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 29 - 5)

(4)螺旋 - 环 - 螺旋(HLH)

螺旋 - 环 - 螺旋(HLH, helix - loop - helix)蛋白具有共同类型的模体:含有两个亲水的 α - 螺旋,链长为 40 ~ 50 aa,两个 α - 螺旋由很长的连接区(环)相连,螺旋的一侧有疏水氨基酸,两条链依赖疏水氨基酸的相互作用形成同二聚体或异二聚体。此螺旋区长约 15 ~ 16aa,含有各种保守的残基,连接环的长度不等,一般为 12 ~ 28aa。与 DNA 结合的是碱性区,长为 15 aa,其中有 6aa 为碱性。在 HLH 中带有碱性区的肽链称为**碱性 HLH**(bHLH)。bHLH 又分为两类。A 类是可以广泛表达的蛋白,包括哺乳动物的 E12/E47(可和免疫球蛋基因增强子中的元件结合)和果蝇 *da*(daughterless,性别控制的总开关基因)的产物;B 类是组织特异性表达的蛋白,包括哺乳动物的 MyoD[肌浆蛋白(myogen)基因的转录因子]和果蝇的 AC - S(achaete-scute, 无刚毛基因的产物)。这种 bHLH 可能是一种转录调控蛋白,大部分以异二聚体形式存在。

(5)同源异型结构域(homeodomains, HD)

同源异型盒(homeobox)是长 180bp 编码 60aa 的序列,几乎存在于所有真核生物中,它的名字是来源于最初在果蝇中鉴别出的同源异型座位(homeotic loci)。它存在于很多控制果蝇早期发育基因中(见 23 章),在高等生物中也发现了相关基序。HD 的 C - 端区域和原核阻遏蛋白的 HTH 同源,但也有几点不同(图 18 - 50):①HD 的 C 端由 3 个 α - 螺旋构成,螺旋 3 结合于大沟,长 17aa;而阻遏蛋白的 HTH 由 5 个 α - 螺旋构成,螺旋 3 长仅 9 个 aa;②原核的 HTH 以二聚体形式与 DNA 结合,靶序列是回文结构,而 HD 是以单体形式与 DNA 结合,靶序列不是回文结构;③HD N 端的臂位于小沟,而 HTH 螺旋 1 的末端与 DNA 的背面接触。

2. 蛋白质和配基的结合

有的调节蛋白并不直接与 DNA 接触，而先和配基结合被活化，如甾类受体蛋白等。当甾类激素等配基进入细胞后，在细胞质中受体与之结合，结合后受体构象发生改变，形成二聚体，成为活化状态，然后进入细胞核。受体识别特殊的保守顺序 GRE 并与之结合，从而活化了其下游的启动子，使糖皮质激素调节基因开始转录（图 18－51）。诱导产生各种相应的蛋白（表 18－10）。甾类受体蛋白一般都具有 3 个不同的功能区（图 18－52）：①N 端区是激活转录所需的区域，不同受体之间此区的同一性仅小于 15%；②DNA 结合及转录活化区，同一性较高为 94% ~ 42%；③C 端的激素结合和二聚体形成区，同一性为 57% － 15%。

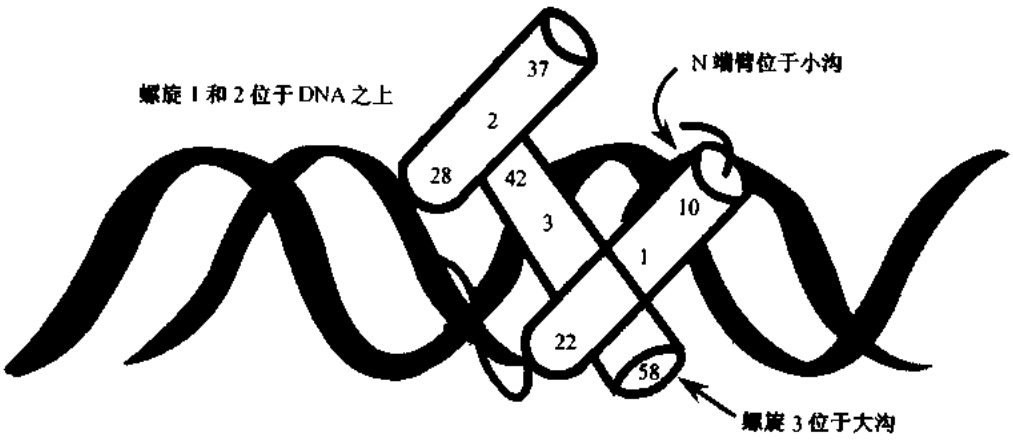


图 18－50　同源异形功能区，螺旋 3 位于大沟。螺旋 1 和 2 位于双螺旋外侧，磷酸骨架及特殊碱基接触、N 端臂位于小沟，产生附加接触。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 29－10）

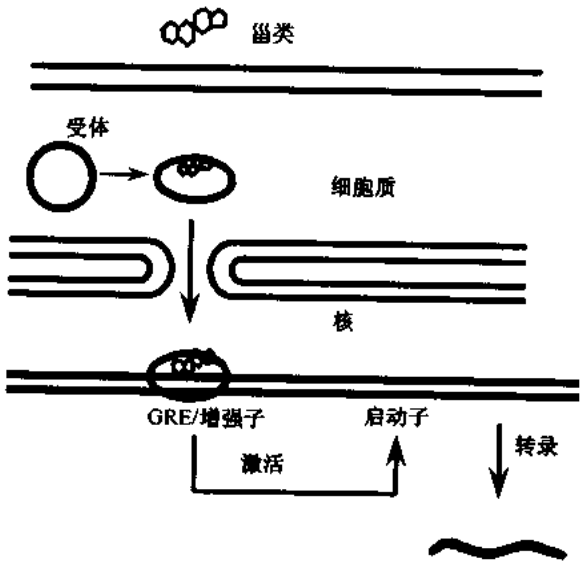


图 18－51　糖皮质激素与受体结合后进入核内和 GRE 结合导致糖皮质激素调节的基因转录。（仿 B. Lewin: Genes Ⅵ, 1994, Fig. 29－7）

3. 蛋白质之间的相互作用

反式的作用因子的另一种调控的形式是蛋白质与蛋白质之间的相互作用。酵母半乳糖代谢中 GAL80 和 GAL4 的相互作用就是很好的例子。酵母半乳糖代谢调控和原核生物 gal 操纵中的调控形式是完全不同的：①被调节的基因绝大部分都是位于不同的染色体上，不连锁；②起负调节作用的蛋白 GAL80 不直接和 DNA 结合，而是通过和作为转录激活因子的 GAL4 蛋白结合，占据其功能域来阻遏转录的；③作为正调节因子不仅需要半乳糖作为诱导物，还需要

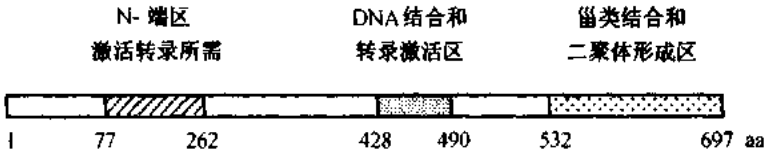


图 18-52 糖皮质激素受体三个功能区的分布。

表 18-10 甾类激素诱导的一些蛋白

激素	组织	诱导蛋白
雄激素	输卵管(母鸡)	伴清蛋白*
		溶菌酶*
		卵清蛋白*
		卵粘蛋白*
	肝(雄鸡)	卵黄蛋白原
		低密度蛋白
		极低密度蛋白
		脂蛋白
	肝(蛙)	卵黄蛋白原
	脑下垂体(大鼠)	促乳素
糖皮质激素	肝(大鼠)	酪氨酸转氨酶
	肾(大鼠)	磷酸烯醇丙酮
		羧基激酶
	脑下垂体(大鼠)	生长激素
黄体酮	输卵管(母鸡)	抗生物素蛋白*
		伴清蛋白*
		溶菌酶*
		卵粘蛋白*
	子宫(大鼠)	子宫珠蛋白
	肝(大鼠)	α-u 珠蛋白

* 卵白蛋白

Peter J. Russell: Genetics, 3rd ed., 1992, Table 17.1

GAL4 蛋白作为转录因子，它不是像原核中的正调节蛋白仅和操纵子中的操纵基因结合，使整个基因簇得以转录，而是分别和各个被调节基因上游元件 UAS 结合，促进转录(图 18-53)。

GAL4 蛋白和 λ 的阻遏蛋白相似，有 DNA 结合功能域，又具有转录活性功能域(图 18-54)，它以二聚体的形式结合于被调节基因上游特殊序列 UAS_G (upstream activator sequence - galactose)上。UAS_G 由 4 个相同的 17bp 序列构成，每个序列都呈两侧对称。若一个序列被结合，那么转录激活是低水平的，若 4 个序列都和 GAL4 结合的话，表达达到最高水平。

GAL4 的结构是利用**功能区转换**(domain-swapping)实验获得的。其 768～881 氨基酸区域为转录激活功能区，可以和转录因子相结合，这个区域中的 851～881 区域又可特异地和 GAL80 蛋白相结合。当半乳糖缺乏时 GAL80 就在此区域和 GAL4 相结合，占据了其转录激活功能区，此时的 GAL4 仍可和 UAS_G 结合，但失去了转录激活能力，因此

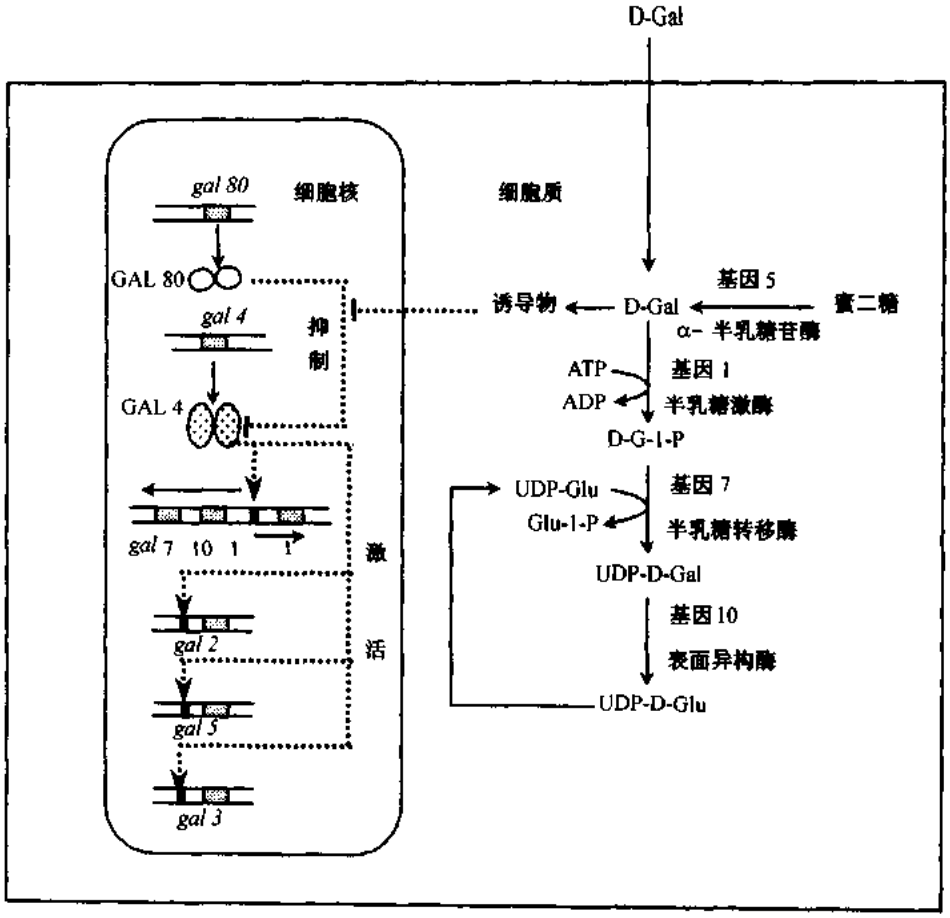


图 18 - 53　酵母半乳糖代谢途径及其调控。图例：----➤：激活；----┤：抑制

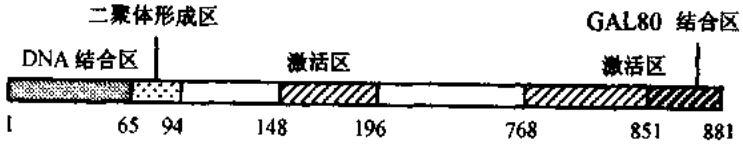


图 18 - 54　GAL4 结合域的分布。

gal 基因不转录；当半乳糖存在时它可以和 GAL80 结合，使其构象发生改变，从 GAL4 上解离下来，这样 GAL4 又可以和转录因子相结合，恢复了转录活性的功能（图 18 - 55）。

前面介绍过的 TF Ⅱ D 和 TF Ⅱ B(见第 12 章)的相互作用也属于蛋白质之间的相互作用调节转录的例子。

4. 蛋白质的修饰

蛋白质的磷酸化和去磷酸化是转录活性调节的一种普遍形式，例如在细胞周期调节中，M 期促发因子 MPF(M-phase-promoting factor 或 mitosis-promoting factor 或 maturation promoting factor)起到了举足轻重的作用。这种蛋白由两部分组成，p34 蛋白是其催化亚基，周期蛋白(cyclin)是其调节亚基。在不同的时期，p34 蛋白的 Thr14Tyt15 Thr161 三个

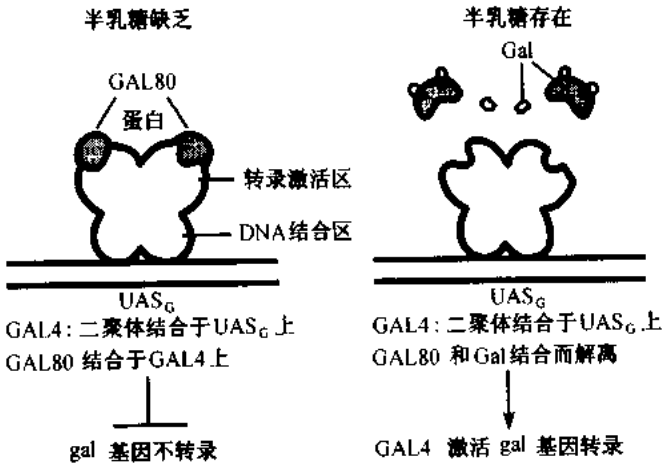


图 18－55 酵母 Ga'基因的激活模型。(仿 pefer J.Russell:Genetics 3rd ed.,Fig.17－7)

位点分别被磷酸化和去磷酸化来调节活性。

MPF 是一种蛋白激酶，激活后可以进一步激活转录因子，使有关蛋白磷酸化，促使 S 期的 DNA 合成和 M 期的染色体凝聚等一系列周期变化。

第六节 转录起始和加工的调节

一、转录起始的选择

在真核生物中同一个基因由于转录起始的不同可以产生不同类型的酶。例如酵母蔗糖酶基因以一种分散的多基因家族存在，有 6 个基因 (*Suc* 1～5, 7) 位于不同的染色体上。每一个基因都可以转录合成蔗糖酶，但有胞内酶和胞外酶两种不同形式。前者的合成不受葡萄糖存在的影响，但含量低；后者的合成受到葡萄糖的抑制。葡萄糖的存在与否使其活力相差 100～1000 倍，两种酶的结构相似，胞外酶仅多了信号序列，经切除后胞外酶比胞内酶在 N 端仅多了一个 Ser。经分析发现 *suc*-2 基因有 3 个 TATA 框，但尚不清楚和 2 种酶的关系，也没有确定是否存在 cAMP－CAP 位点(图 18－56)。

小鼠的唾腺、肝脏和胰脏都能合成 α－淀粉酶，但在 3 个组织中 α－淀粉酶的浓度不同。小鼠的第 3 号染色体上有两个连锁的 α－淀粉酶基因 *amy*-1 和 *amy*-2，*amy*-1 在唾腺和肝中表达，*amy*-2 却在胰脏中表达。在唾腺中产物的浓度是肝中的 100 倍。这是由于在不同的组织中使用了 *amy* 基因 5′端的 2 个不同启动子。在唾腺中使用的启动子 P_S 较强，转录活性比 P_L 高 30 多倍，但唾腺细胞中 *amy* 的 mRNA 浓度要比肝脏中的高 100 倍，表明可能还受到其他调控因素的影响。

二、选择性加工

即使是同一个基因，相同的初始 mRNA，但由于 5′端，内含子及 3′末端等选择的不同，使成熟的 mRNA 也不同，结果编码了功能不同的蛋白。

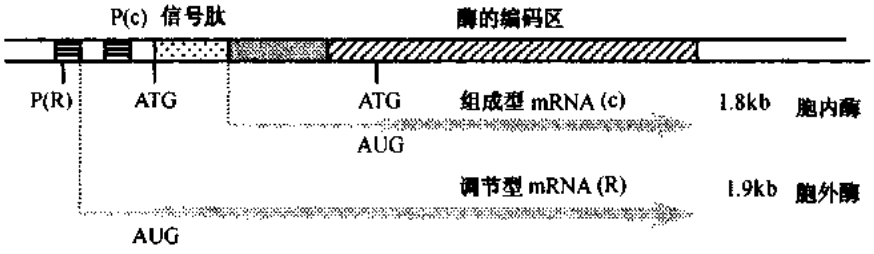


图 18－56　酵母 No9 染色体上的 *suc－2* 基因的不同转录起始。

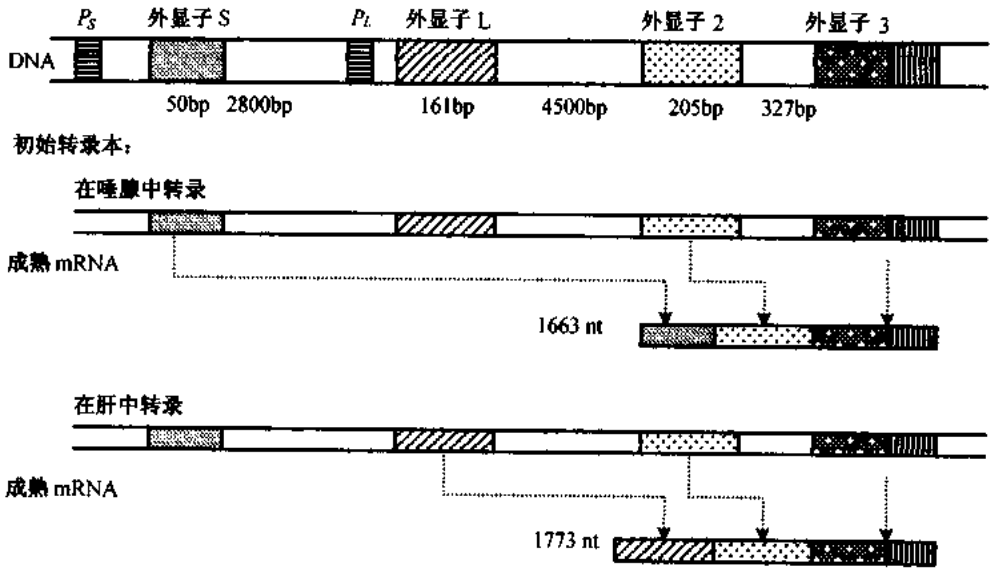


图 18－57　在小鼠中，淀粉酶(*amy*)基因利用不同启动子产生两个不同的 mRNA。
(仿 B. Lewin: Genes Ⅳ, 1990, Fig. 25－5)

1. 不同 5′端的选择

前面所举的小鼠淀粉酶的例子实际上也是属于 5′端的不同选择。这是由于一个基因具有两个启动子区，每一区都有它自己的组织调控元件，那么两个长度不同的转录本将会产生组织特异性 mRNA。例如鸡的肌球蛋白(myosin)轻链基因在心脏和砂囊中转录后产生的成熟 mRNA 就不同，前者为 LC1(light chain)，后者为 LC3，它们具有相同的 3′编码区，但 5′编码区不同(图 18－58)，在大鼠中也发现编码的肝球蛋白链的单个基因在不同组织通过不同的转录后加工来调节表达。

2. 选择不同的 3′端

同样的一个基因在不同组织中由于 3′端加尾位点的选择不同也可产生不同的 mRNA，而形成不同的产物。如大鼠甲状腺中合成的降钙素(calcitonin)和脑下垂体合成的神经肽(neuropeptide)，都是由同一个基因编码的，由于 3′端加尾位点的选择不同，使其 mRNA 的 3′端的编码区不同，导致最终合成的产物也完全不同(图 18－59)。

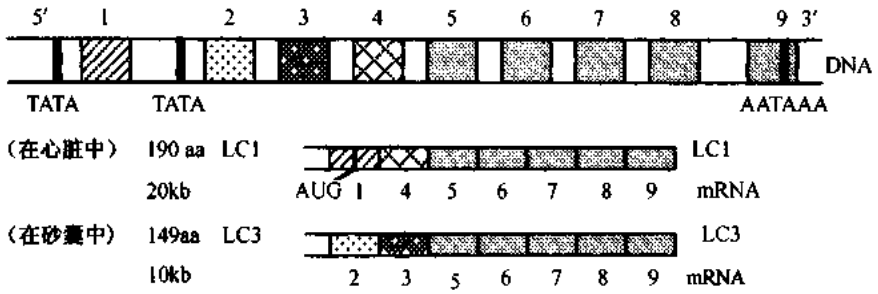


图 18－58 鸡肌球蛋白轻链前体 mRNA 在不同组织中进行不同的选择性拼接。

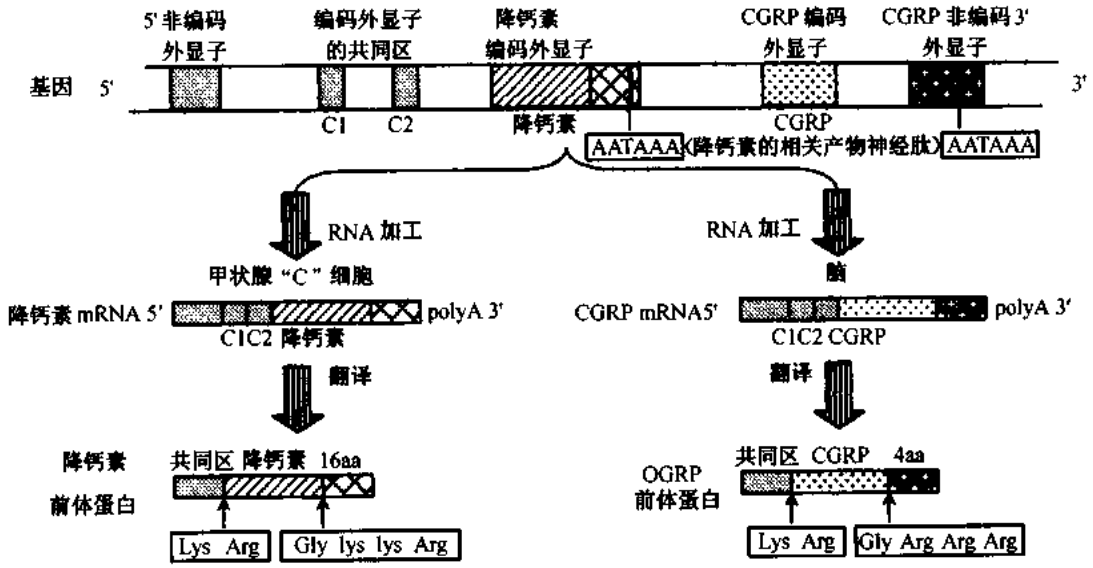


图 18－59 大鼠降钙素基因两个 polyA 位点和不同剪接模式。(参考 James D. Watson *et al.*: Molecular Biology of The Gene, 4th ed., 1987, Fig. 21－37)

3. 选择不同外显子

例如大鼠的肌钙蛋白(torponin T)基因在不同的发育阶段以及不同横纹肌种类中由于选择性剪接内含子,结果产生了不同的肌钙蛋白 T。

第七节 翻译的调控

原核生物基因表达的调控主要在转录水平上进行。真核生物由于 RNA 较为稳定,所以除了存在转录水平的调控以外,在翻译水平上也进行多种形式的调控。

一、mRNA 运输控制

运输控制 (transport control) 是对转录本从细胞核运送到细胞质中的数量进行调节。真核和原核生物不同,有一个核膜包被的核,核膜就是基因表达的一个控制点。

我们知道初始转录本是在核内被广泛地加工。实验表明几乎有一半的蛋白编码基因的初始转录本一直留在核里面，然后被降解掉。成熟的 mRNA 如何从核内转运到细胞质中呢？看来这些 mRNA 都要通过核孔进行转运，但是对于从核中输出的过程以及输出或保留所需的信号知道得很少。某些证据表明 snRNPs 对于 mRNA 留在核中是很重要的。例如在抑制剪接体装配的成熟酵母中，mRNA 易于从核中输出。人们提出了**剪接体滞留模型**(spliceosome retentior model)。在这个模型中剪接体的装配与 mRNA 的输出相竞争，当前体 mRNA 在剪接体加工的过程中，RNA 滞留在核中，不能与核孔相互作用。当加工完成后，内含子被切除了，mRNA 从剪接体上解离下来，游离的 mRNA 能与核相互作用，但内含子不行。现在还不清楚 mRNA 是否需要一个特殊的输出信号还是属于无规则的输出。

二、mRNA 翻译的控制

通过核糖体对 mRNA 分子的选择充当了翻译调节的主角。不同的翻译明显地影响到基因的表达。例如 mRNA 储存在很多脊椎和无脊椎动物的未受精卵中，在未受精阶段蛋白质合成率很低，但一旦受精蛋白质合成立即增加。因此这种合成的增加并非由于新的 mRNA 的合成，可能是由于存在一种翻译控制之故。最近认为这种翻译控制主要是蛋白降解控制，在控制中蛋白降解的速率是受到调节的。

在细胞质中所有的 RNA 都要受到**降解控制** (degradation control)，RNA 降解的速率(也称为 RNA 的转换率)是受到调节的。通常核糖体中的 rRNA 和 tRNA 是很稳定的，相比之下 mRNA 分子的稳定性很不一致，有的 mRNA 的寿命可延续好几个月，有的只有几分钟。我们在某些类型的细胞中加入调节物可使某些特殊蛋白的合成增加。这可能涉及到相关基因转录速率的增加，也可能涉及到其 mRNA 稳定性的增加。表 18－11 表明某些特殊效应分子对各种组织和细胞中的 mRNA 稳定性的影响。

mRNA 的降解可能是基因表达调控的一个重要控制点，mRNA 降解速率的不同和各 mRNA 结构特点有关。特别是 mRNA 的选择性降解在很大程度上是由于核酸酶和 mRNA 内部结构相互作用的结果。例如在很多短寿命的 mRNA 3′端非翻译区(UTR)中的一组富含 AU 的序列 (UUAAUUUAU) 是和它们的不稳定性有关系的，但现在还不清楚 AU 丰富区怎样使 mRNA 不稳定的，可能和去除 poly(A)有关；也可能 AU 序列与 80S 复合物形成过程中的某种因子结合。

表 18－11　特殊效应分子对于各种组织和细胞中的 mRNA 稳定性的调节

mRMA	组织或细胞	调控因子	mRNA 的半衰期	
			有调节因子	无调节因子
卵黄蛋白原	肝(蛙)	雌激素	500h	16h
卵黄蛋白原	肝(母鸡)	雌激素	~24h	<3h
极低密度脂蛋白原	肝(母鸡)	雌激素	~20－24h	<3h
卵清蛋白/伴清蛋白	卵巢(母鸡)	雌激素，孕激素类	>24h	2－5h
酪蛋白	乳腺(大鼠)	促乳素	92h	5h
前列腺甾类结合蛋白	前列腺(大鼠)	雄激素	增加 30 倍	

Peter J. Russell: Genetics, 3rd ed., 1992, Table 17.2

三、mRNA 的结构和翻译的效率

5′m^7G 帽结构是否存在和是否易于接近 eIF－4F 对翻译效率有着明显的影响。起始密码子 AUG 的位置和其侧翼的序列对翻译的效率也有影响，这些因素主要是通过与调控蛋白、核糖体、RNA 等的亲和性改变影响到起始复合物的形成，并影响到翻译的效率。

5′端非翻译区的长度也会影响到翻译的效率和起始的精确性，当此区长度在 17～80nt 之间时，体外翻译效率与其长度成正比。此区的高级结构和高 G-C 含量对翻译的起始都有妨碍。

5′端非翻译区的二级结构影响到调控蛋白与帽结构的接近，阻碍 40S 前起始复合体的装配和在 mRNA 上的扫描，起负调控的作用。但若二极结构位于 AUG 的近下游，(最佳距离为 14 nt)，将会使移动的 40 亚基停靠在 AUG 位点，增强起始反应。真核的系列翻译起始因子可使二级结构解链，使翻译复合体顺利通过原二级结构区，继续其肽链的延伸，而不会起阻碍作用。在这种情况下二极结构又起到了正调控的作用。

MRNA 3′端的 poly(A) 不仅和 mRNA 穿越核膜的能力有关，而且影响到 mRNA 的稳定性和翻译效率。有 ploy(A) 的 mRNA 翻译效率明显高于无 poly(A) 的 mRNA ，Poly A 长度和翻译效率有关。有人将 poly(A) 比做翻译的计数器，随着翻译次数的增加，poly(A) 在逐步缩短，也就是说 poly(A) 越长 mRNA 作为模板的使用的半衰期越长。Poly(A) 对翻译的促进作用需要 **PABP**[poly(A) 结合蛋白] 的存在，PAPB 结合 poly(A) 最短的长度为 12 nt，当 poly(A) 缺乏 PABP 的结合时，mRNA 3′端的裸露易招致降解。PABP 迁移到 AU 序列时，导致 poly(A) 的暴露促进了 mRNA 的降解。

四、翻译的起始调节

免网织红细胞和大多数生物的网织红细胞一样是没有细胞核的，因此没有 DNA，而 mRNA 也早已加工好了，所以蛋白质的合成调节只能依赖于翻译水平，很多基因表达水平的资料就是从网织红细胞的研究中获得的。血红素对珠蛋白合成的调控就是一例，这种调节是通过对翻译起始的复合物的形成来控制的。

在网织红细胞中有两个合成酶系，一个合成血红素，另一个合成珠蛋白。血红素通过自身的反馈调节来控制 (图 18－61)，而其浓度又可调节珠蛋白合成的速率，使珠蛋白合成速度为血红素的两倍，这种调控是通过控制翻译起始复合物的形成来进行的。

依赖 cAMP 的蛋白激酶 (R_2C_2) 是由调节亚基 R_2 和催化亚基 C_2 构成，当它和 cAMP 结合释放出自由的 C 亚基，成为活化的蛋白激酶，而血红素的存在对这一步反应是抑制的。自由的 C 亚基可催化无活性的 **HCR**(控制血红素阻遏物 hemin-controlled repressor，又称 eIF－2 激酶) 磷酸化而成有活性的 HCR，而有活性的 HCR 又使 eIF－2 磷酶化失去活性。而血红素抑制这一系列反应就可使 eIF－2 不被磷酸化，保持活性状态 (图 18－61)，和细胞中的 tRNAini 及 ESP(eIF-2-stimulating protein) 形成翻译起始复合物，开始合成珠蛋白，因此缺乏血红素时，珠蛋白也不能合成。

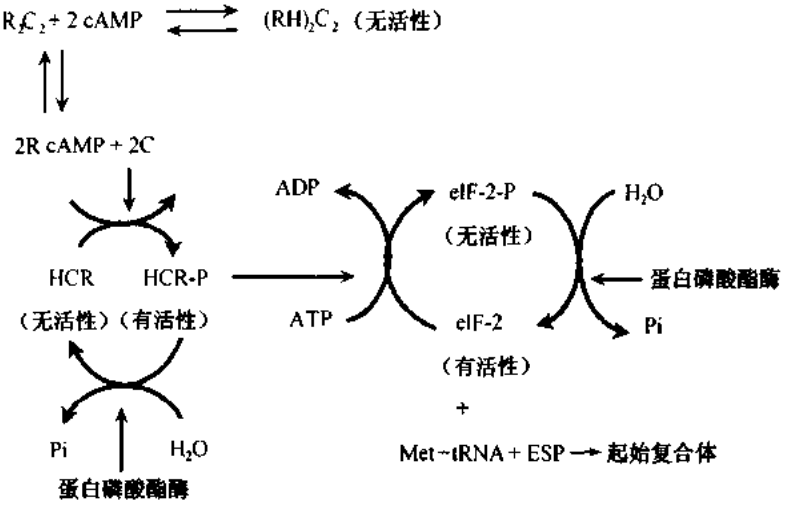


图 18－60　血红素对网织红细胞中蛋白质合成的调节。HCR：控制血红素阻遏物(hemin－controlled repressor)；ESP：延伸起始因子－2 刺激蛋白(eIF－2－stimulating protein)。

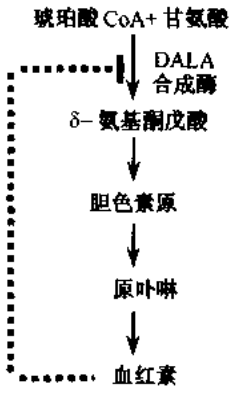


图 18－61　血红素的反馈调节。DALA：δ－aminolevulinate(δ－氨基酮戊酸)。

五、选择性翻译

在原核生物中常通过操纵子来控制合成的相关蛋白浓度比，而在真核中不存在操纵子，所以就要采用别的方式，选择性翻译就是其中一例，比如 α 和 β 珠蛋白的合成。众所周知珠蛋白是由两条 α 链和两条 β 链组成的。但在二倍体细胞中都有 4 个 α－珠蛋白基因，如果它们转录和翻译速率相同的话它们之间的浓度比应是 α：β＝2：1，而实际上是 1：1，那么是通过转录调控还是通过翻译调控使它们的产物达到以了平衡呢？人们进行了以下的体外实验，在无细胞系统中加入等量的 α-mRNA 和 β-mRNA 以及少量的起始因子，结果合成的 α-珠蛋白仅占 3%，说明 β-mRNA 和起始因子的亲和性远大于 α-mRNA。当加入过量的起始因子时 α 珠蛋白和 β 珠蛋白之比为 1.4：1，接近 1：1。表明在翻译水平上存在的差异，即和翻译起始因子的亲和性不同使产物达到平衡。这种差异是由于 mRNA 本身的二级结构和高级结构所致。

六、反义 RNA 调控

在原核生物反义 RNA 发现以后，人们开始注意真核生物中反义 RNA 的存在及功能。1984 年 Adelman 等发现大鼠的促性腺激素释放激素(GnRH)的基因两条链都能转录，首次在真核中发现了反义 RNA；1986 年 Green 等发现来自骨髓细胞瘤病毒的癌基因 *myc* 三个外显子中的第 1，2 两个外显子之间有部分互补。在有的细胞中，当失去外显子 1 时，*myc* 基因过量表达，推测外显子 1 可能通过互补来抑制 *myc* 的表达。现已弄清在 *C. elegans* 中，控制幼虫发育的基因 *lin 14* 受到 *lin 4* 基因的反义调节（见 16、24 章）。这是通过对翻译模板的抑制来进行调控的途径。人们已将反义 RNA 发展成一门

反义技术应用于动、植物病毒的抑制，果蔬的保鲜，甚至应用于肿瘤的治疗。反义药物的开发无疑具有着广泛的前景。

七、翻译的自我调节

在真核生物中也存蛋白质合成的自我调节。例如微管蛋白是构成纺锤体的主要成分。而秋水仙碱和长春花碱都能抑制维管蛋白的多聚化，从而使细胞中游离的微管蛋白浓度增加。若在组织培养细胞中加入秋水仙碱或长春花碱侧会使微管蛋白的mRNA消失，如果用微量注射器将微管蛋白注入到哺乳动物细胞中，那么就会抑制微管蛋白的进一步合成。其产生的机制和模型我们已在16章中作了介绍。可能过量的微管蛋白结合于核糖体的新生蛋白上或结合于mRNA上，阻止翻译，导致mRNA的降解所致。

第八节　细胞周期的调控

真核生物细胞有丝分裂周期有两个关键的过程：①DNA的复制；②姐妹染色单体均匀地分配到子细胞中。这两个过程的进行要求精确而协调。细胞怎样控制在S期中起始和终止DNA的复制呢？怎样控制染色体在M期产生凝聚呢？显然细胞必须具备控制细胞周期的机制。

细胞周期的调控与个体发育，分化，生长，再生，衰老以及细胞的癌变都密切相关，因此细胞周期调控近年来已成为分子生物学研究的热点之一。从20世纪70年代开始细胞生物学家，遗传学家和发育生物学家在各自的领域中以不同的途径对细胞周期调控进行了深入研究，共同提示出周期调控的框架和诸多调控因子互相作用所形成的极联调控网络，其核心因子便是p34蛋白和周期蛋白(cyclin)。

一、细胞周期与调节

用酵母突变体以及爪蟾早期胚胎来研究细胞分裂周期结果表明：在细胞周期中有两个主要的限制点，第一个称"**起点**"(start)，位于G期末(图18-62)。在此刻，DNA合成起始，不久就进入S期。第二个是M期进入限制点(commitment to M phase)。位于M期的开始，此时染色体开始凝聚，最终姐妹染色单体发生分离。细胞能通过这两个限制点关键就在于调节蛋白提供了某种信号。

早在1970年Johnson等将分裂期细胞与间期细胞融合，结果间期细胞出现了早熟染色体凝聚现象。人们意识到分裂期细胞中有一些因子可以诱导间期细胞进入分裂期。卵母细胞中的这种因子被称为**成熟促进因子**(muturior promoting factor，MPF)。后又称为**M期促发因子**(M-phase promoting factor，MPF)。现已正实MPF的普遍存在于从酵母到人类所有真核细胞的M期细胞中。与此同时遗传学家们筛选了大量酵母的突变株来研究细胞周期的调控。20世纪70年代初Hartwell等根据多种温度突变株的遗传分析，查明了30个与细胞周期有关的基因。Nurse等长期对裂殖酵母进行分析，提示了*cdc2*及其他基因产物组成的调控系统调节酵母M期的启动。发育生物学家们在20世纪80年代初进行一系列出色的研究，在海胆卵裂的过程中发现了周期蛋白，后在海星，爪蟾，果蝇，酵母等生物中也发现了周期蛋

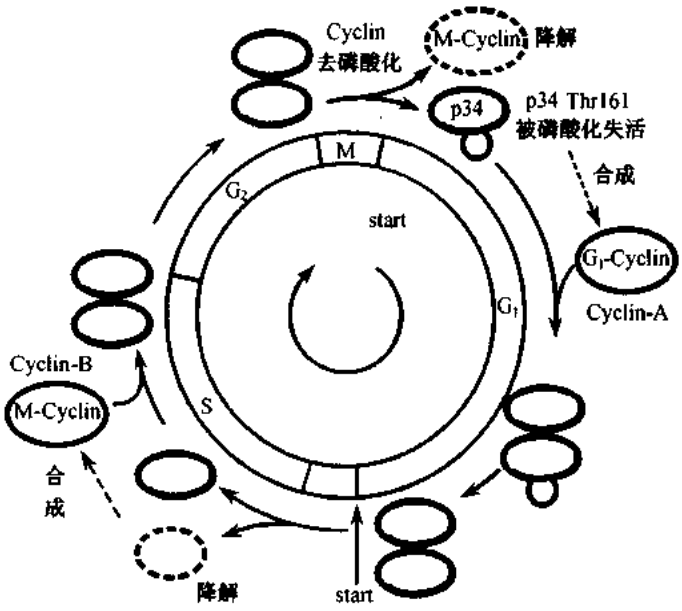


图 18－62　细胞周期的调节。(摘自 Murray A.W.Nature 1989,342:14)

白。

1988 年 Lohka 等从成熟的非爪蟾卵母细胞提取 MPF 并进行了高度纯化,发现它由两种成分组成,分别为 45kDa 和 32kDa,他们立即用裂殖酵母 p34 抗体对纯化的 MPF 进行免疫印迹和免疫沉淀实验，结果证明此 32kDa 蛋白是 p34 蛋白的同源物。对 45kDa 蛋白进行测序发现它就是在海胆等生物中发现的周期蛋白。

周期蛋白的功能是作为 MPF 的调节亚基，在 G_1 和 G_2 期积累，在 M 期和 $S \rightarrow G_2$ 过程中降解。p34 分子量为 34kDa，是一种“Start”限制点和 M 期的特异蛋白磷酸酶，它有 3 个磷酸化位点：Thr14,Tyr15,Thr161，这些位点在不同的时期，不同的位点进行磷酸化来激活一系列细胞周期有关的蛋白。其中 Tyr15 和 Thr161 位点为主，p34 蛋白 Thr161 位点磷酸化使 MPF 具有活性，而 Tyr15 位点的磷酸化又可抑制 Thr161 位点的磷酸化。如 Thr14 和 Tyr15 在 G_2 期磷酸化 MPF 是没有活性的，在进入 M 期时必须去磷酸化，Tyr161 在进入 M 期时必须被磷酸化。在粟酒裂殖酵母中 p34 蛋白由 *cdc2* 基因编码，在酿酒酵母(*S.cerevisiae*)中 p34 蛋白由 *CDC28* 基因编码。其功能是 MPF 的催化亚基，在细胞周期中含量稳定，只有磷酸化和去磷酸化状态的变化，与周期蛋白相互作用来调节 DNA 的复制起始和 M 期的进入。p34 蛋白在不同时期和两种周期蛋白结合，一种是 M－周期蛋白，也叫做 B－型周期蛋白，另一种 G_1 周期蛋白，也叫做 A－型周期蛋白，p34－cyclin A 和 p34－cyclin B 功能相似，但识别的底物不同，前者控制 M 期的进入，后者控制 Start 限制点的通过。显然 cyclin 和 p34 蛋白激酶是周期调控的关键成分，它们的同源蛋白已在多种真核生物(人、蛙、海胆和海星等)中被发现(表 18－12)。

p34 蛋白属于 Ser/Thr 型的激酶,但只有与 cyclin 结合后才具有活性。

二、细胞周期的调控

M 期激酶(p34－cyclin B)的催化活性是怎样激活的呢? p34 是一个磷酸化蛋白,而

表 18－12 几种生物中细胞周期的周期调控同源蛋白

物种	催化亚基(p34 蛋白)		调节亚基(周期蛋白)	
	M→G_1	G_1→M	M→G_1	G_1→M
粟酒裂殖酵母	cdc2	cdc2	cdc13(B-like)	cig1、2
酿酒酵母	CDC28	CDC28	CLB 1～4(B-like)	CLN 1～3
哺乳动物/蛙	p34	cdk2(p33),cdk4	Cyclins A、B1、B2	Cyclins A、D1、D2、D3、E

参考 B. Lewin: Genes Ⅵ, 1997, Fig36－18

且它的磷酸化状态决定了它的活性。P34－cyclin B 的活性化要求在某些位点 (Tyr15) 去磷酸化，而另一些位点(Thr161)磷酸化。

在小鼠细胞中，p34－cyclin 必须在 p34－的两个 ATP 结合位点去磷酸化，那就是 Thr－14 和 Tyr－15，此反应需要一种特殊类型的磷酸化酶 (能作用这两个磷酸化位点)，或者两种不同磷酸化酶(cdc25)分别从两个残基上去磷酸化。有可能 p34－cyclin B 是自我激活的，这种激酶少量的活化就能触动其余部分完全活化。

另一个磷酸化位置是在 p34 的 Thr－161。它是在 G_2 期被磷酸化，在 M 期末去磷酸

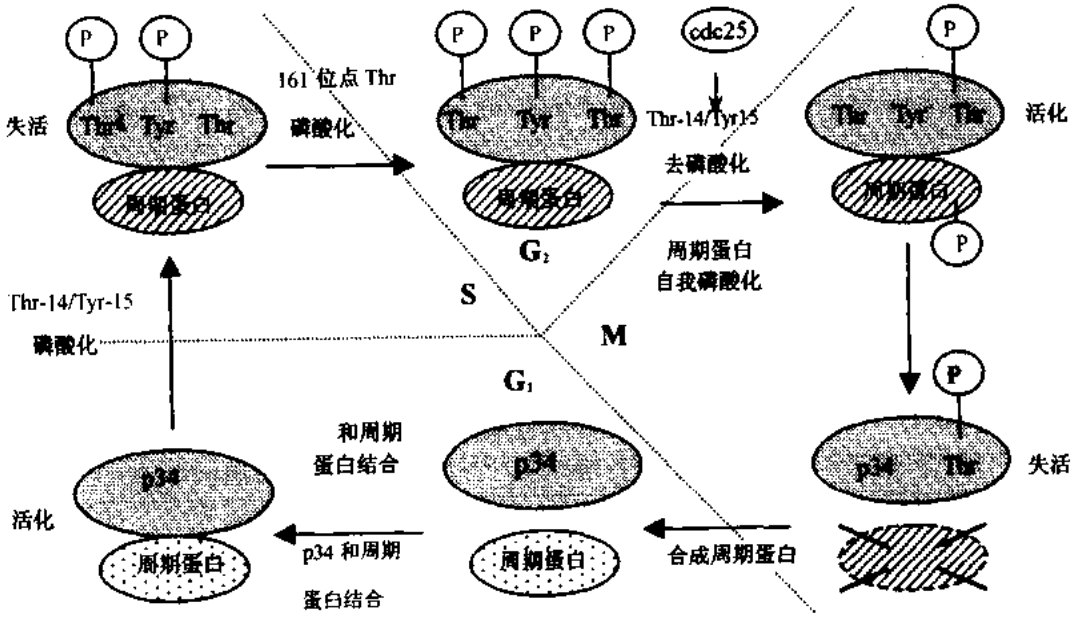


图 18－63 细胞周期的调控。(参考 B. Lewin: Genes Ⅵ, 1997, Fig. 36－7)

化。此位点的磷酸化对于 p34 的活化是需要的，此位点发生突变的 p34 就不能磷酸化，从而也失去了活性。

按理说 cyclin 与一个失活的 p34 结合将会产生一个无活性的激酶，但事实令人吃惊，在体外 cyclin B 可特异地与 Tyr 去磷酸化的 p34 结合，并产生一个有潜在活性的二聚体。但二聚体的形成又导致了 Thr－14/Tyr－15 被磷酸化。这样 p34 和 cyclin B 结合增强活性。二聚体的活性一直维持到去磷酸化才失活。

在 M 期 cyclin 亚基的降解，使 p34－cyclin B 失活，实际上这是细胞结束 M 期的需要。切除 N－端的 cyclin B 对蛋白酶是有抗性的。在爪蟾卵中合成缺失 N 端的 cyclin。

或者在游离的细胞中提取出 cyclin 都会使细胞分别停止在中期。这表明 p34 - cyclin 激酶活性的丧失是完成有丝分裂的前提。

一个重要的问题是 cyclin A 和 cyclin B 在有丝分裂中如何发挥不同的作用，关键是在不同的时间合成和降解，但尚无直接的证据表明在细胞中两种 cyclin 对于通过有丝分裂是否都需要，在不同的细胞类型中是否需要不同的 cyclin。

cyclin 的稳定性也是动物细胞周期调节的一个环节，这种稳定性和 CSF（cytosstatic factor 细胞静止因子）有关。CSF 是爪蟾中的一种因子，它可使卵母细胞停止在减数分裂Ⅱ期。MFP 可诱导细胞进入减数分裂或者有丝分裂期，但并不能使分裂停止，而 CSF 可使细胞停止在 M 期。这是由于 CSF 可使 MPF 活性稳定，在 M 期不至失活，结果 M 期就不能继续进行。

原癌基因 *c-mos* 的产物是一种 Ser/The 激酶，可活化 CSF，CSF 的降解可使细胞 M 期继续进行。CSF 有可能是一种 cyclin B2，在体外实验中，cyciln B2 是 *c-mos* 游酶的一种底物。看来 CSF 可以稳定或维持 MPF 的活性。另一种可能 c-mos 的底物是蛋白酶，它可降解 cyclin，当它被磷酸化而失活后，cyclin 就不被降解。但 cyclin B 并不能使 G_1→S 期。

在 M 期活化的 p34 - cyclin B 可以使下面一组蛋白磷酸化。① 使层连蛋白（lamins）磷酸化，使得细胞在分裂期核膜解体；②使组蛋白 H1 磷酸化，导致染色体的凝聚；③使微管蛋白磷酸化，导致细胞骨架的重新组合，为细胞分裂做好准备；④使泛蛋白连接酶体系（Uoiguitin ligase system）磷酸化，它们是依赖泛素的蛋白水解酶，它们经磷酸化激活后可使 cyclin - B 的降解，促进细胞进入 G_0/G_1，（表 18 - 13）此作用正好和 CSF 的作用相反。

在 G_1 期 cyclin B 被降解后，被磷酸化的 p34 失活，相应地基因开始合成 cyclin A（图 18 - 63），其半衰期较短，边合成，边降解，其合成的量大于降解的量，一旦合成的产物达到一定阈值，然后和 p34 结合形成 p34 - cyclin A，p34 的 Ser217 被磷酸化，p34 - cyclin A 活化，具有丝/苏氨酸蛋白激酶活性，能激活某些转录因子，使这些因子

表 18 - 13　在 M 期活化的 p34 可能使一些靶蛋白磷酸化而激活

p34 的靶蛋白	作用时间	作用
H1 组蛋白	M 期	M 期染色体凝聚
层连蛋白（lamins）	M 期早期	核膜破裂
核仁素（nucleolin）	M 期早期	停止核糖体合成
肌球蛋白轻链	M 期	抑制胞质分裂
p60src	M 期	进一步磷酸化靶蛋白
SW15	G 和 M 期	阻遏核的聚集
微管蛋白	M 期后期	细胞骨架的重组
泛蛋白连接酶体系	M 期后期	使 Cyclin-B 降解

能与细胞周期盒序列（ACGCGTNA）相结合，起动与 DNA 复制有关的基因群：*cdc6*, *cdc8*, *cdc9*, *cdc21*, 它们分别编码 DNA 合成酶，胸苷激酶，DNA 连接酶和胸苷酸合成酶，使 DNA 得以复制。同时具有活性的 p34 还可激活转录因子 Oct1，使其结合到组蛋白 H2B 的启动子区，促进表达，使细胞进入 S 期。

在细胞从 S 期进入 G_2 期的过程中 cyclin A 降解，p34 失活。在 S 期相关的基因合成了 cyclin B（图 18 - 62）与 p34 结合形成 MPF，*wee1* 或 *mik1* 基因的产物是一种蛋白磷

酸激酶，可使 cdc2 磷酸化而失活，但 Nim1 基因的产物又可 Weel 基因产物磷酸化而失活，起负调节作用；相反 Bws1 和 SDS21 基因产物又可使 Weel 的产物去磷酸化而活化。另外的激酶可进一步使 p34 的 Thr161 磷酸化，这对于 p34 的激活是重要的，但由于 Thr14、Thy15 仍处于磷酸化的状态，所以这时的 p34 - cyclin B 还没有活性。当一些有丝分裂的因素如有丝分裂原（mitogen）存在时可以活化 cdc25 基因的产物，此产物也是一种蛋白磷酸酶，可特异地使 p34 的 Thr14/Thr15 去磷酸化，结果导致了 M 期激酶（p34 - cyclin B）的活化，结果周期从 G_2 进入 M 期。

在真核细胞周期中，核内 DNA 的复制（S 期）和有丝分裂（M 期）是有联系的，必须完成 DNA 的复制然后细胞分裂才能开始，为了使这一过程的连接精确而有序，必须存在某种系统，它可以发现未复制的 DNA 并产生抑制信号，阻止有丝分裂因子的激活。近期发现一种 DNA 结合蛋白 - 染色体凝聚调节蛋白（RCC1，regulator of chromosome condensation），在其缺失时有丝分裂可在 DNA 合成前启动。这就意味着 RCC1 可能是一种信号分子，它能发现未复制的 DNA，并产生细胞分裂的抑制信号。但 RCC1 的突变体不仅不能调节细胞周期，同时也失去了一些其他功能，表明 RCC1 在核中有多方面的功能。

第十九章

癌基因与癌

基本概念

癌细胞和肿瘤细胞不同，具有无限生长、转化和转移三大特点。

确立细胞系与正常细胞不同，但还不是癌。它可以无限分裂，但仍保持了生长控制：表现为培养的贴壁特点，血清依赖性，密度制约抑制和细胞骨架的形成。

癌基因又叫转化基因，是人类和其他动物基因组中的一类基因，常担负着正常的生理功能，一旦发生突变。无论突变位点是在上游调控区还是在编码区都会导致肿瘤的形成。

人和动物细胞中的癌基因称为细胞癌基因。在病毒基因组中癌基因的同源序列称为病毒癌基因。具有正常生理功能的癌基因又称为原癌基因。现已发现100多种癌基因。

抗癌基因也称为肿瘤抑制物，现已发现10种抗癌基因，其中研究得最为透彻的是Rb和p53基因。

癌基因的致癌作用常和细胞内信号传递有关，抗癌基因的作用常和细胞周期的调控有关。

癌的发生是十分复杂的。癌基因的突变或抗癌基因的失活是基本的原因。

第一节　肿瘤细胞的特点

所有高等真核生物都有一定的寿命，其体细胞的生长和分裂也是受到高度调节的，但癌细胞却不是如此，它们失去了生长的控制。当一个正常细胞变成肿瘤时在三方面发生明显的改变。

（1）无限生长（immortalization），这和生长控制发生改变有关。

（2）转化（transformation），失去了生长的正常抑制，例如不受生长因子的支配。

（3）转移（metastasis），癌细胞具有侵袭正常组织的能力，可以从原来的组织转移到其他的组织，在那儿建立新的克隆。

正常的细胞也可以变成肿瘤细胞。从脊椎动物中取得的细胞进行培养，它们可以传50代左右，然后进入**极限**（crisis）（图19－1），培养的细胞大部分死亡，而少数幸存的细胞发生了某种变化，使其能突破限制无限地分裂。其分子机制尚不清楚，但近年来认为可能和端粒酶的复活有关。Harley（1990）提出端粒的长短可反映细胞有丝分裂的次数。0～93岁不同年龄人的成纤维细胞其端粒的长度和细胞分裂有关。早老症患者的细胞端粒明显比正常人短，其细胞增殖能力降低。他还认为正常人体细胞的端粒达到一定长度时即开始短缩，并启动终止细胞分裂的信号，细胞便开始衰老死亡，他称之为**第一死亡期**（M1）。抗癌基因 *P53* 和 *Rb* 的突变都能使细胞逃逸M1，获得一定的额外增殖能力。

细胞的寿命达到极限会死亡，此称为**第二死亡期**（M2）。越过M2生存下来的细胞有无限增殖的能力，且端粒酶为阳性，癌细胞是无限增殖的，故提出癌和端粒酶的活性有关。端粒酶的复活可延长细胞的寿命，肿瘤细胞的无限增殖的特征可能是端粒酶异常复活的结果。肿瘤细胞的端粒较短，但端粒酶具有活性，可能是端粒酶被激活以稳

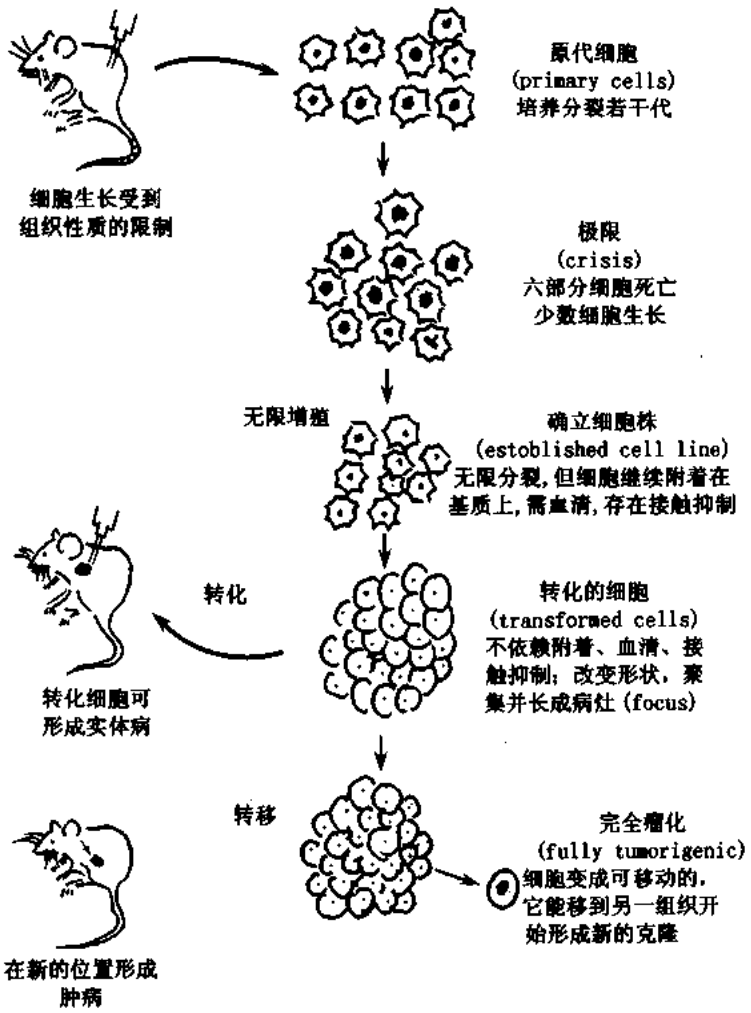


图 19-1 从正常细胞变成癌细胞的步骤及产生的三大特点:无限增殖、转化和转移。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 37-1)

定被降解的末端。

原代细胞(primary cells)是从生物体直接取出的细胞它们表现出了体内的表型特征。但在大部分情况下只能存活相对较短的一段时间,到了极限时培养细胞都会死掉。超过此极限的一些细胞建立形成了一种非致瘤的确立细胞株(Establised cell line),它们能无限制地存活,且某些特点发生了改变,此可能是在适应培养的过程中发生的。这些改变可能和肿瘤形成的过程相同。

确立细胞系虽然无限分裂,但一般并不致癌,仍保持了生长控制,表现的特点有一部分仍和原代细胞相似,如:①**贴壁依赖性**(anchorage dependence)细胞附着在固体基质或薄膜的表面。②**血清(或生长因子)依赖性**(serum or growth factor dependence):血清可提供一些必要的生长因子。③**密度制约抑制**(density-dependent inhibition):细胞生长要求一定密度范围。细胞彼此的接触可抑制细胞的生长。④**细胞骨架的形成**(cytoskeletal organization)。细胞是扁平的,而且沿着附着基质的表面延伸。细胞内部由肌动蛋白构成了较紧的纤维网。

以上表明细胞仅仅在固体基质表面上生长,形成很薄的单层细胞。这些特点提供了一些参量,根据这些参量可以详述细胞的正常状态。当然任何确定细胞株不能和体

内的状态完全等同，因此仅提供一种体内控制的近似值。用这种细胞株来分析生长控制的遗传背景必须要谨慎，因为在培养中染色体的成分常会发生改变，或者不是真正的二倍体，可能是非整倍体。

肿瘤细胞培养中发生各种性质的变化，其第二步变化称为**转化**，一个转化的细胞的生长几乎不受什么限制。通常不需要固体表面供其附着，因此细胞相互聚拢而不再延展开。同时减少了对血清的依赖性，形成厚厚的一层细胞团，叫做**转化灶**(focus)，如将其接种到实验动物体中可导致肿瘤的产生。

将正常细胞和肿瘤细胞进行比较可以帮助我们了解肿瘤形成的遗传基础，也能弄清和转变有关的表型变化。

各种化学试剂增加了细胞转化的频率，它们被称为**致癌物**(carcinogenic)。这些致癌物分为两类，一类是"**起始**"(initiate)肿瘤的形成，另一类是"**促进**"(promote)肿瘤的形成。表明肿瘤的发生存在着不同阶段。

直到最近我们还没有完全弄清癌的分子基础，但我们已知道存在两种基因，即**癌基因**(oncogene)和**抗癌基因**(antioncogene)，它们的结构或产量发生改变都会导致癌的产生。

第二节　癌基因和抗癌基因

一、癌基因

癌基因(oncogene)又叫**转化基因**(transforming gene)是人类或其他动物基因组中含有的一类基因，它们一旦激活便能导致正常细胞发生癌变。人和动物细胞中的癌基因被称为**细胞癌基因**(cellular oncogene，c-*onc*)，未活化的 c-*onc* 称为**原癌基因**(proto-oncogene)。病毒基因组中癌基因的同源序列称为**病毒癌基因**(virus oncogene，v-*onc*)。

早在 1911 年就发现将患白血病家禽的细胞提取物注入正常家禽体内可引起白血病，且发现肿瘤细胞中含有病毒，P.Rous 于 1911 年发现了一种病毒可使鸟类产生肿瘤，此病毒命名为 **Rous 肉瘤病毒**(Rous sarcoma virus)。由于这项工作的深远意义，于 1966 年他荣获了诺贝尔奖。1969 年美国的 Huebner 和 Hodaro 提出了癌基因假说，认为所有的细胞中都含有致癌病毒的全部遗传信息，这些遗传信息代代相传，其中与致癌有关的信息称为癌基因。在通常情况下癌基因处于被阻遏状态，只有当细胞内有关的调节机机制遭到破坏的情况下癌基因才表达，而导致细胞发生癌变。

1970 年 H.Temin 发现致癌的 RNA 病毒中存在着一种反转录酶（1975 年他因此获得了诺贝尔奖），于是提出了原病毒假设，认为 RNA 病毒通过反向转录和正向转录以及与宿主细胞 DNA 发生交换或重组，能形成癌基因。

1976 年 H.E.Varmus 和 J.M.Bishop 发现鸟类肉瘤病毒中含有致癌基因（1989 年获诺贝尔奖）。后人发现上述病毒进入细胞后通过反向转录整合到宿主的染色体中形成相应的前病毒，这些病毒和前病毒的基因产物可以致癌。到 20 世纪 80 年代初，随着 DNA 体外重组和动物细胞转化技术的发展通过很多实验证实了癌基因的存在。

Bishop 分离鸟类肉瘤病毒的癌基因后，用分子杂交的方法证明宿主细胞也存在与病毒癌基因相同的 DNA 序列，这就是细胞癌基因。显然反转录病毒和动物的基因

组中都含有相同的癌基因，那么这些癌基因最初是起源于病毒还是起源于动物呢？根据以下几点证据人们认为病毒癌基因是通过转导作用从宿主细胞原癌基因中掳获过去的，即癌基因起源于动物。

(1) 各种动物的 c-*onc* 其数目和位置比较固定，而反转录病毒中的 v-*onc* 数目和位置是不固定的。

(2) 线虫(*C. elegans*)和果蝇等动物并不被反转病毒所感染，但这些动物的基因组中也具有与高等动物癌基因同源的基因；

(3) 在多数高等动物的 c-*onc* 中发现有内含子，而反转录病毒的癌基因(v-*onc*)则无内含子，表明它们似乎已经通过了转录后加工，已不是初始的模板。

(4) 动物的 c-*onc* 为反转病录病毒所转导。

c-*onc* 在细胞的增殖，分化和胚胎发育中都具有重要的功能，不仅高等动物中存在 c-*onc*，低等动物如线虫、果蝇乃至真菌酵母中也存在。只有当 *c-onc* 本身或其调控因子发生突变或畸变时才和肿瘤的形成有关。现已鉴定出约 100 种癌基因(表 19－1)，并将它们分为几组，从跨膜蛋白到核内的转录因子(图 19－2)。它们所行使的不同功能可能有助于我们对肿瘤发生机制的探索。

表 19－1 主要癌基因的名称及性质

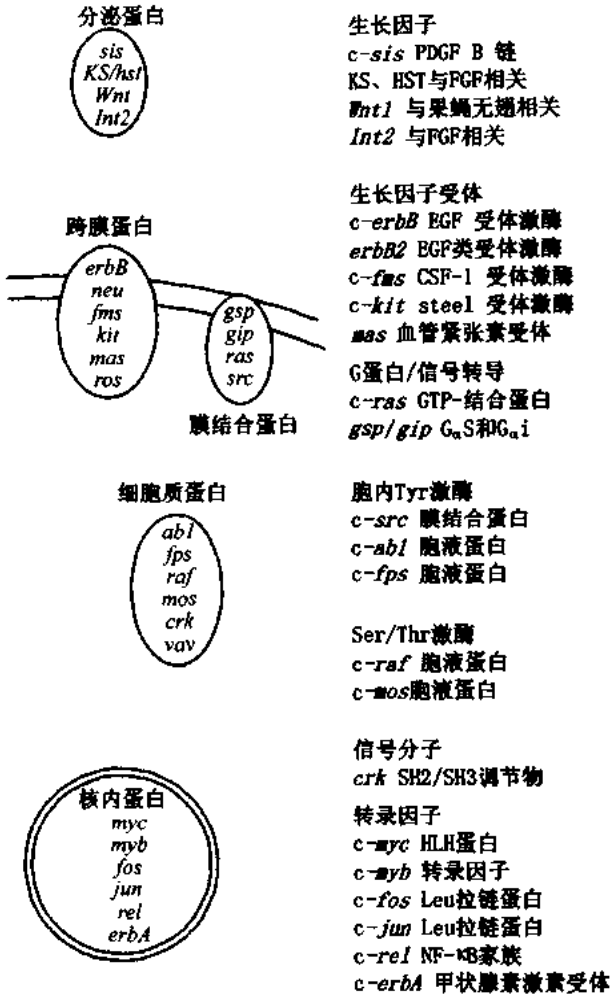
onc	相关反转录病毒	v-*onc* 来源	onc 蛋白	引起的疾病	产物活性	产物定位	基因定位
src	RSV	鸡	$pp60^{src}$	肉瘤	PK(Tyr)	质膜内面	20q12－q13
H-*ras*	Ha-MuSV	大鼠	$pp21^{ras}$	肉瘤，红白血病	G 蛋白	质膜	11p13
K-*ras*	Ki-MuSV	大鼠	$pp21^{ras}$	肉瘤，红白血病	G 蛋白	质膜	6p32－q13
mos	Mo-MuSV	小鼠	$p37^{env-mos}$	肉瘤	PK(ser, thr)	胞质	8q22
fos	FBJ-MuSV	小鼠	$pp55^{fos}$	骨肉瘤	亮氨酸拉链蛋白	核	14q21－q31
sis	SSV	猴	$p28^{env-s}$	肉瘤	与 PDGFβ 链同源	分泌	22q11－qter
sis	Pl-FeSV	猫	$p76^{gag-sis}$	肉瘤	与 PDGFβ 链同源	分泌	22q11－qter
fms	SM-FeSV	猫	$gp180^{gag-fms}$	纤维肉瘤	CSF－1 受体激酶	跨膜蛋白	5q34
fes	ST-FeSV	猫	$p85^{gag-fes}$	纤维肉瘤	PK(Tyr)	质膜	15q12－q25
jun	ASV-17	鸡	$p39^{c-jun}$	纤维肉瘤	Leu 拉链蛋白	核	?
fps	FuSV	鸡	$p130^{gag-fps}$	肉瘤	TK(Tyr)	胞质	15q12－q25
myc	MC29	鸡	$p110^{gag-myc}$	瘤，肉瘤，髓细胞瘤	HLH 蛋白	核	8q24
abl	MuLV	小鼠	$p90－p160^{gag-abl}$	B－细胞淋巴瘤	TK(Tyr)	胞质	9q34－qter
rel	REV-T	火鸡	$p64^{rel}$	淋巴白血病	NF-κB 家族	核	?
erbB	AVE	鸡	$gp65^{erbB}$	红白血病和纤维瘤	EGF－受体 K	跨膜	7pter－q22
erbA	AVE	鸡	$p75^{gag-erbA}$	红白血病和纤维瘤	甲状腺素受体	核	17p11－q21
myb	AMV)	鸡	$p45^{myb}$	髓细胞瘤	转录因子	核	6q22－q24

RSV(Rous sarcoma), Ha-MuSV(Harvey murine sarcoma), Ki-MuSV(Kirsten murine sarcoma) Mo-MuSV(FBJ murine osteosarcoma), SSV(Simian sarcoma), Pl-FeSV(Feline sarcoma), SM-FeSV(Felinesarcoma), ST-FeSV(Feline sarcoma), ASV-17(Avian sarcoma), FuSV(Fujinami sarcoma), MC29(Avian sarcoma), MC29(Avian sarcoma), MuLV(Abelson leukemia), REV-T(Reticuloendotheiosis), AVE(Avian erythroblastosis), AVE(Avian erythroblastosis)

二、抗癌基因

抗癌基因(antioncogene)也称为肿瘤抑制物。现在共发现了 10 种抗癌基因，其中研

究得较为深入的是人的 Rb(retinoblastoma,视网膜母细胞瘤)和 p53 蛋白。



PDGF:platelet-derived growith factor 血小板衍生生长因子
FGF:fibroblast growith factor 成纤维细胞生长因子
EGE:epidermal growith factor 表皮生长因子
CSF:cytostatic factor 细胞静止因子
Steel:青灰因子,癌基因的产物
图 19-2　癌基因编码的各种蛋白。(仿 B.Lewin:Genes Ⅵ,1997,Fig.37-12)

Kundson 早在 1971 年根据对遗传性和散发性视网膜母细胞瘤患者的调查研究,提出著名的"二次突变假设(two mutation hypothesis)"。认为视网膜母细胞变成瘤细胞必须要经过二次突变。第一次突变存在或发生于上一代的生殖细胞中,即患儿的体细胞已有了一次遗传下来的突变,只要在视网膜母细胞内发生第二次突变就可致癌。后来遗传学家发现家族性患儿的体细胞中 13 号染色体存在缺失(13q14),表明该区域可能存在与视网膜母细胞瘤发生有关的基因 *RB*。接着 Sparkes(1980) 将 *RB* 基因定位于 13q14。Benedict(1983)提出位于 13q14 的 *RB* 的一对等位基因均失活才会产生该肿瘤。表明 *RB* 是以隐性方式起作用。而与开始认为 *RB* 为显性遗传的观点不同。现已证实 *RB* 的作用确为隐性的。Stephe 等(1986)开始构建了人的 13 号染色体的 λ 噬菌体文库,其中 1.5kb 的 DNA 片段和 Rb 有关。接着又分离到一个 30kb 的 cDNA 片段,用此片段

可检测到 13q14 带 70kb 区域的基因。这对分离和克隆 *RB* 是至关重要的。同年眼科专家 Dryja 发现 *RB* 失活时可以致瘤，杂合子(RB^+/RB^-)易被诱变而失活，这完全符合二次突变学说和隐性遗传的假设。他还发现经眼球摘除手术能存活下来的患儿长大后易患成骨肉瘤，乳腺癌和小细胞肺癌，当时这是一个谜。现在已知和 *RB* 广泛的抑癌作用有关。1986 年 Friend 将 13q14 位点的 *RB* 经 cDNA 克隆制成各种探针，检查很多例骨肉瘤和组织肉瘤患者发现约 50% 的患者 *RB* 基因纯合缺失。这意味着 *RB* 具有抑制的作用。为了证实这个问题 1987 年加州大学欧文分校的 Stanbridge 等分离了人类 13 号染色体，将它注入培养的成骨肉瘤细胞中，结果导致了癌细胞逆转，同年加州大学圣地亚哥分校的李文华把 *RB* 转化到培养的癌细胞中使其逆转。更进一步证明了 *RB* 的抑癌作用。他克隆了 *RB* 全长的 cDNA，发现其产物是一种长 928 氨基酸，分子量为 105kDa 的蛋白质。

Rb 的分子功能是什么呢？它主要是可以和其他各种蛋白相互作用，其中包括肿瘤抗原：SV40 T 抗原，腺病毒 E1A 抗原，人类乳头瘤病毒 E7 抗原，可结合到 Rb 的 379～792 残基区域。一种可能是部分致癌蛋白由于和 RB 产物的相互作用其致癌的作用受到了抑制。另外 Rb 还有抑制细胞增殖的作用。

Rb 是一种 105kb 的核内磷蛋白，它对细胞周期也有一定的影响。保持在 G_0/G_1 期的静止期细胞中 Rb 是未磷酸化的，在细胞周期中它被周期蛋白 CDK(cyclin－CDK) 复合物磷酸化。大部分情况是发生在 G_1 期末，一直延伸到 G_2 期末，在有些分裂中是去磷酸化。Rb 在非磷酸化状态才能和几种蛋白如 SV40 T 抗原等特异结合。这些相互作用只发生在细胞周期的特定阶段(在 S 期前)。Rb 经磷酸化就会释放这些蛋白。

靶蛋白中包括一组 E2F 转录因子，它们所激活转录的靶基因其产物是 S 期所必需的。和 Rb 结合以后 E2F 的激活转录的能力受到抑制，表明 Rb 可以阻遏那些依赖 E2F 的基因的表达。通过这种途径 Rb 可直接阻止细胞进入 S 期。而 Rb-E2F 的复合体也可直接阻遏某些靶基因。因此 E2F 从 Rb 上解离以后这些基因才能表达(图 19－3)。

一些病毒的肿瘤抗原特异地与非磷酸化形式的 Rb 结合，研究得最多的是 SV40 T 抗原和腺病毒的 E1A。图 19－3 表面明 Rb 在正常细胞中与 E2F 结合，在转化细胞中与肿瘤抗原结合的模型。非磷酸化的 Rb 可以阻止细胞的扩增。为了要使细胞周期得以进行必须要阻遏 Rb 的活性。其途径就是周期蛋白使 Rb 磷酸化。当肿瘤蛋白和非磷酸化的 Rb 结合时也能抑制其活性。由于 Rb－肿瘤复合物不再和 E2F 结合，使 E2F 得以以游离状态进入 S 期，活化 S 期需要的一些基因。

RB 的过量表达妨碍细胞生长。丢失了 Rb 骨肉瘤细胞株的特点就表明了 *RB* 对于抑制细胞扩增的重要性。当将 *RB* 转导到此细胞株中其生长就被抑制了。然而这种抑制作用又可被 D-cyclin 的表达所解除。这是由于 D-cyclin 所形成的 CDK-cyclin 复合体可使 RB 磷酸化。p107 和 p103 也是和 Rb 同一类型的蛋白，它们具有相关的序列和相同的特点。

图 19－4 表示细胞周期和肿瘤发生之间的关系。几种细胞周期的调节蛋白在肿瘤中被发现是突变失活的，因而被认为是肿瘤的抑制物。除了 Rb 以外还有一些小的抑制蛋白(最引人注意的 p16 和可能性大的 p21)及 D-cyclin。虽然这些蛋白(最著名的是 RB)在增殖细胞周期中所起的作用与致瘤有关，这大部分是在静止(G_0 期)细胞中发现的。在静止细胞中 Rb 是未被磷酸化，D-cyclin 的水平较低或者缺失，p16，p21 和 p27 蛋白使 cdk-cyclin 失活。

p21(cip1/ wAF1)，p27(Kip1) 和 p15/ p16(ink4) 合称为 Cki 蛋白，它们结合在

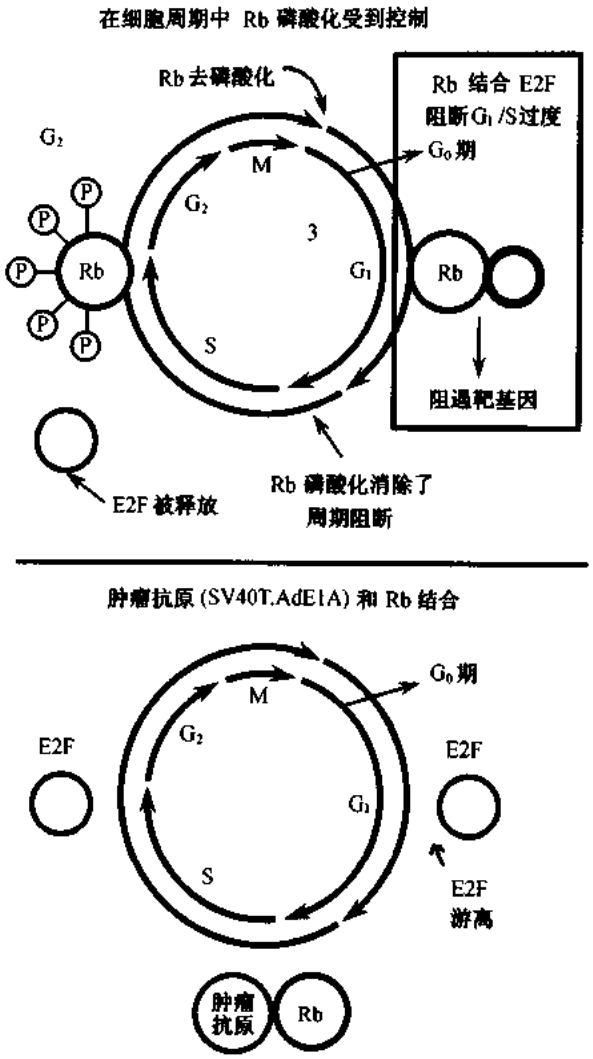


图 19－3　当 Rb 被磷酸化(在正常细胞中)或和肿瘤抗原结合时(在转化细胞中)细胞周期阻断被消除。(仿 B.Lewin:Genes Ⅵ,1997)

cdk－cyclin 二聚体上,使其失活。如在静止期通过这种途径阻止 Rb 的磷酸化,使细胞周期不能进入 S 期。

p15/ p16 由两个相邻的基因编码，二者密切相关。p16 可特异地结合于 cdk4 和 cdk6,表明它们与 G_0/G_1 期的过相关。p16 不能抑制丢失了 Rb 的细胞的增殖,表明它的功能是阻止 cdk-cyclin 使 Rb 磷酸化,p16 通过和 cdk 亚基的结合还可抑制 cdk4-cyclin D 及 cdk6-cyclin D 的装配,也可能抑制了它们的激酶活性。

p21 是一种通用的 cdk 抑制物,结合所有的 cdk2、4、6 复合体,表明它是通过 G_1/S 各阶段来阻断过渡。前期培养细胞(直接取自动物的细胞)中的 cdk-cyclin 二聚体常以一种四聚体的形式存在,其实是结合了另外 2 种蛋白:一是 PCNA,是真核 DNA 聚合酶 δ 的成分(见第 11 章),和复制有关;另一个则是抑制物 p21。在转化细胞中 cdk-cyclin 复合体丢失了 p21 和 PCNA,表明 p21 可能涉及到 G_1/S 过渡的控制。

p27 蛋白有一段序列和 p21 部分同源。也能混杂地和 cdk－cyclin 复合物相结合。p27 的过量表达可阻断 S 期的过渡。当细胞用 TGF-β(来处理，细胞进入静止期时 p27 的水平增加。p21 和 p27 阻断 cdk-cyclin 二聚体的催化亚基被 CAK(cdc2－activation ki-

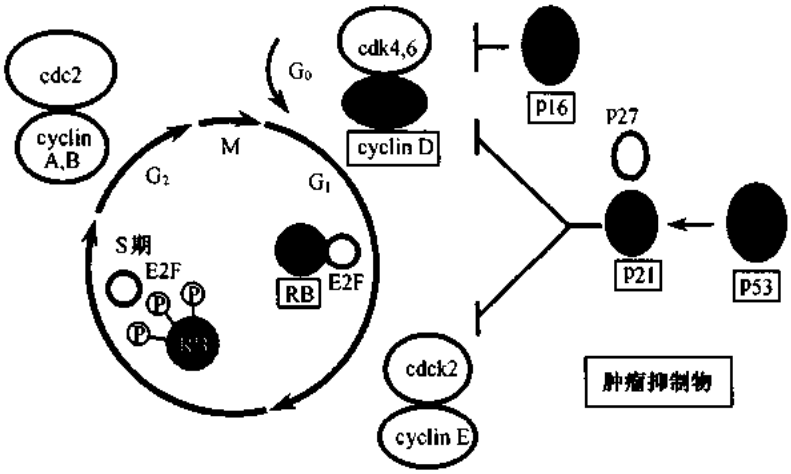


图 19－4　几种和 G_0/G_1 或 G_1/S 周期控制有关的几种蛋白是肿瘤抑制物。
（仿 B.Lewin:Genes Ⅵ,1997,Fig.37－21）

nase)磷酸化而激活。它们也阻止 cdk-cyclin 复合体的催化活性(图 19－5)。

P53 基因是一种肿瘤的抑制物，但在 Li-Fraumeni 综合征中是作为常染色体显性遗传的，这种综合征是一种罕见的遗传性癌，患者可在不同组织中出现癌。此是由于 *P53* 基因发生错义突变。由于突变使 p53 的结构改变，产生了新的特性，这就解释了长期以

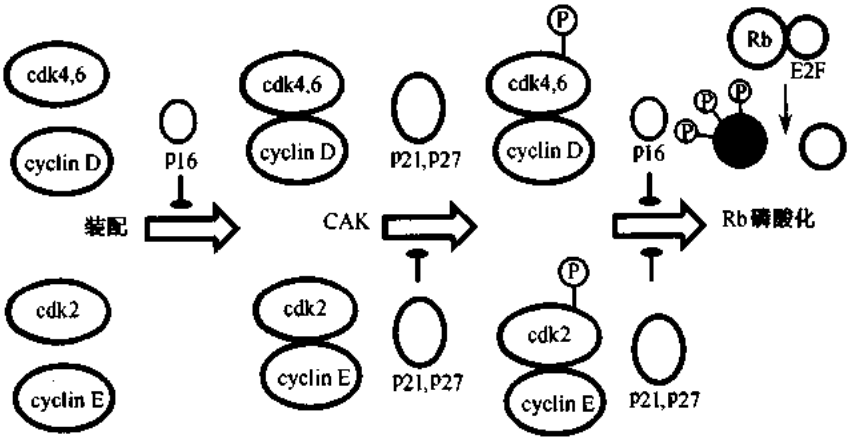


图 19－5　P16 抑制 cdk4,6 和 cyclin D 的装配及 cdk4,6－cyclin D 对 Rb 的磷酸化；p21,p27 抑制 CAK 对 cdk－cyclin 复合体的磷酸化及此复合体对 Rb 的磷酸化。
（仿 B.Lewin:Genes Ⅵ,1997,Fig.36－24）

来发现它和致癌有关的一些现象。

p53 也是一种核内的磷蛋白，是最主要的肿瘤抑制物，人类一半以上的癌都显示出 p53 蛋白的缺失或是 p53 基因的突变。它开始是在 SV40－转化细胞中发现与 T 抗原相结合。在很多转化细胞或肿瘤细胞系中发现 p53 蛋白的量大大增加。在早期实验中发现克隆的 p53 会导致细胞无限制生长，因此把 p53 作为一种癌基因，通常它的作用是显性的。

所有的 p53 的转化形式都是此蛋白的突变形式。它们成为一种显性失活突变体，其功能掩盖了野生型蛋白的功能阻止其发挥作用。显性失活突变大部分情况是形成一种异聚体蛋白，它含有突变型亚基和野生型亚基。在异聚体中野生型亚基是没有功能的。p53 可能以四聚体的形式存在。当 p53 突变型亚基和野生型亚基结合时，此四聚体

呈现突变体构象。图 19－6 表明野生型的 *P53* 基因产物可以限制细胞的生长。若缺失了一对等位基因，或者有一个基因发生显性实变，失去了限制的功能，细胞可以无限生长。

从分子水平来看 p53 的功能是什么呢？此有两种学说。①p53 被发现时是以 SV40 病毒的大 T 肿瘤抗原的结合蛋白存在的，它能与 T 抗原结合这是野生型 p53 蛋白的特点，而突变型 p53 是不能和 T 抗原结合的。野生型 p53 与 T 抗原的结合干扰了 T 抗原在 SV40 DNA 复制中启动作用。可能是野生型 p53 与细胞中某些 T 抗原的类似物结合，抑制了它们的活性，但在 p53 突变体中就失去这种功能。②p53 也是一种 DNA 结合蛋白，它能识别长 10bp 两侧对称的模体，并激活含多拷贝这种基序的启动子的转录。在有的座位上它可能阻遏靶基因。p53 通过调节基因的表达来控制细胞周期，缺乏这种调节就会导致细胞无限生长。

大部分 p53 的错义突变位点在很多生物中都是保守的。表明突变可能干扰了某种保守的功能。p53 突变体产生了一些新的特点包括：①其半衰期从 20 分钟增加到几小

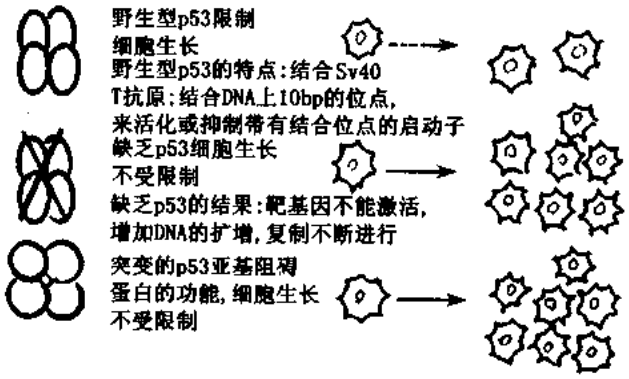


图 19－6　野生型 p53 限制细胞的生长，一对野生型等位基因的缺失或一个等位基因产生显性突变都可使细胞生长不受限制。
（仿 B. Lewin：Genes Ⅵ，1997，Fig. 37－21）

时；②构象发生改变，可用其抗体来测定；③改变了从核到细胞质的定位；④不再和 SV40 T 抗原结合；⑤不再和 DNA 10bp 的模体结合。

在所有正常的细胞中 p53 的水平是低的。但当 DNA 通过照光或其他处理而发生损伤时 p53 被激活，导致 p53 大量增加，p53 的活化可以引发两类效应。细胞停止生长或凋亡直至死亡。图 19－7 表明细胞在 G_1 的早期 p53 触发细胞周期的限制点，阻断了细胞周期的进程，这是通过诱导 p21 对 CDK-cyclin 激酶的抑制来实现的，使损伤的 DNA 在细胞进入 S 期前得到修复的机会。但若细胞已进入分裂期，那么 p53 就触动细胞死亡的程序。这种凋亡（apoptosis）的典型结果是细胞坍缩成一个固缩的小团以及核 DNA 断裂。

p53 有多种分子活性（图 10－8）

（1）p53 是一种 DNA 结合蛋白，可识别 10bp 的回文模体，第 120～290 位氨基酸的区域负责与特定靶序列的结合。

（2）可激活含多拷贝模体启动子转录。在其 N 端第 1～42 位氨基酸区域承担此激活转录活性。p53 还可以阻遏某些不含有以上模体的基因，其机制不明。

（3）p53 也能和损伤的 DNA 结合（第 310～390 位），C 端的功能区可以识别单链

DNA 区。

(4) p53 是一个四聚体,其寡聚化对于其突变体具有的显性失活调控的作用是先决条件,若不形成四聚体,那么一个分子的突变就难以影响到整个四聚体构象的改变而失去功能。

当 p53 作为转录因子时用它的第 120～290 位功能域与靶序列结合。N 端的激活功能域直接与 TBP(TATA box－binding protein)相互作用。此区域也是其他相互作用蛋白的靶位点,包括细胞蛋白 mdm 2 和腺病毒的癌蛋白 E1B(55kDa)。mdm 2 是一种细胞蛋白,它通过和 p53 N 端结合,抑制其活性,阻断转录。p53 诱导 mdm 2 的转录,这样它们之间的相互作用形成了负反馈,限制了彼此的活性。

和 E1B 55kDa 蛋白的相互作用使得腺病毒阻断 p53 转录活性。p53 其他的区域也可以成为抑制作用的靶位点;SV40 抗原结合到 p53 特异 DNA 结合区,这样阻止了它识别靶基因。

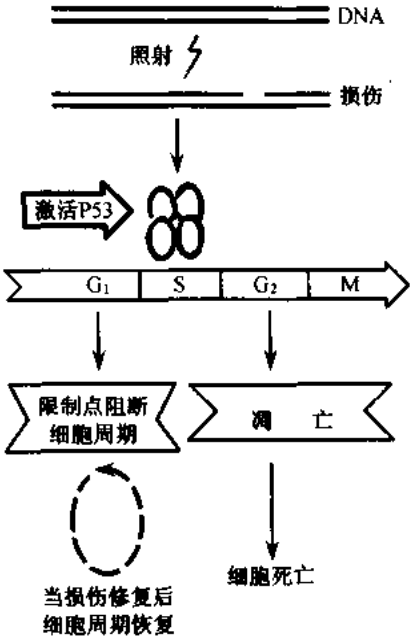


图 19－7 DNA 的损伤激活 *P53* 这个结果取决于细胞周期处于哪一阶段。在细胞周期的早期 *P53* 激活限制点。阻止细胞周期前进直到损伤修复为止。如处于很晚的阶段 *P53* 引发细胞的萎缩。(仿 B. Lewin: Genes Ⅵ, 1997, 37－23)

p53 的功能是作为一种转录因子,激活一系列的基因,包括 p21(抑制细胞周期,只执行 G_1 限制点所必要的), GADD45(一种修复蛋白,它也可通过对照射损伤作出反应而被激活);这两个基因都涉及到细胞周期的制止。从这两个基因和其他靶基因的激活,看来似乎意味着 p53 使生长停止。

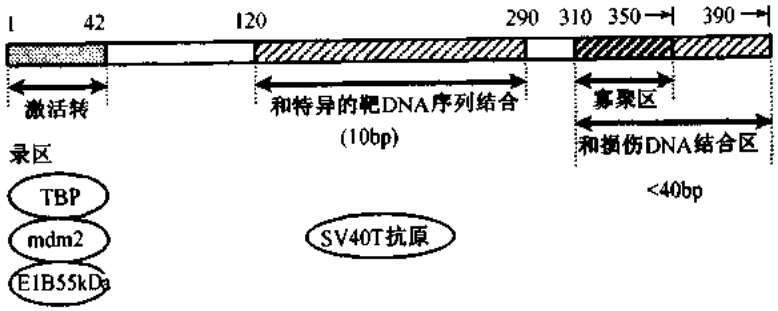


图 19－8 p53 蛋白不同的结构功能域。(参考 B. Lewin: Genes Ⅵ, 1997, Fig. 37－24)

p53 突变的细胞有新增加的特点,即 DNA 扩增,可能导致癌细胞中所发现的基因组不稳定的特点。

p53 的 C 端可以和非特异的短单链 DNA 区(＜40nt)相结合或和由于很短的(1～3b)缺失及插入而产生的错配区域相结合。这样的靶位点是由 DNA 的损伤产生的。相互作用的结果激活了特定结构域的特异序列结合活性,从而 p53 刺激靶基因的转录。这种连接的特点并不清楚。可能要经过二个阶段。当 p53 通过 C－端功能区和 DNA 损伤位

点结合时，其性质发生改变，然后它与损伤位点解离，结合到靶基因上，将其激活。

p53 触发细胞凋亡的能力尚缺乏了解，这与转录靶基因的能力之间不存在相关性。腺病毒 E1B(19kDa)蛋白可阻断 p53 凋亡细胞的活性，但并不阻断其激活靶基因的活性。p53 独立于制止生长之外的凋亡细胞的这一功能是通过 E1B 55kDa 蛋白能阻断转录活性而不影响凋亡而引起人们的关注。有可能 p53 通过别的途径来触发凋亡，而不涉及靶基因的激活。

肿瘤的发生和凋亡的缺乏二者之间关系的重要性是通过 *bcl2 onc* 显示出来的。*bcl2* 原来被鉴定为一种靶基因，它在某些肿瘤中通过易位而被激活，已清楚它具有抑制各种凋亡途径的特点。表明凋亡在抑制肿瘤形成中起到一种重要的作用。可能由于它消除了潜在的致瘤细胞。当 *bcl2* 被激活阻止了凋亡，使这些细胞免于死亡。

带有 p53 功能缺陷的细胞具有多种表型；这种多效性使得决定它的这些效应与肿瘤抑制功能之间的直接联系很困难，关于 p53 作用的大部分知识是来自于其被激活的情形。我们推测它触发制止生长或凋亡的作用和它抑制肿瘤的能力有关。很显然 p53 对 DNA 损伤若缺乏反应将会增加细胞对致癌性突变的易感性。但我们不知道这是否是 p53 惟一的作用。我们不知道在正常情况下当 p53 未被 DNA 损伤激活时它可能起到什么作用。$p53^-$ 小鼠能够正常的存活表明 p53 的作用对于发育来说并不是必要的。

Rb 和 p53 的特点表明二者是肿瘤的抑制物，它们以某种方式控制细胞的增殖，若缺乏二者就会失去这种控制而导致肿瘤的形成。这两种蛋白都具有与细胞周期相关的功能，当它们为活性状态时可抑制肿瘤，失去它们细胞就会无限生长。

第三节　癌基因产物的结构、功能和细胞癌变

在图 19－9 中已表明 c-*onc* 产物有各种正常的功能，当其结构和产量发生改变都会引起细胞癌变，那么具体的机制是什么呢？现在我们举几个例子来说明这个问题。

一、*src* 基因

src 的产物 p60*src* 具有 TPK(酪氨酸蛋白激酶)活性，分子量约为 60kDa，与细胞的增殖和转化有关。Src 蛋白有四个结构域(图 19－10)：①N 端十四烷基化位点是和膜结合的位点；②SH2(*src* homology region-2) 结构域能和自身的或其他蛋白的磷酸化酪氨酸位点结合，传递信号。③416 位 Tyr 的自身磷酸化和细胞的转化活性有关（图 19－11)。④527 位 Tyr 的磷酸化(图 19－10)，当受体无活性时，Src 527－Tyr 自身磷酸化并

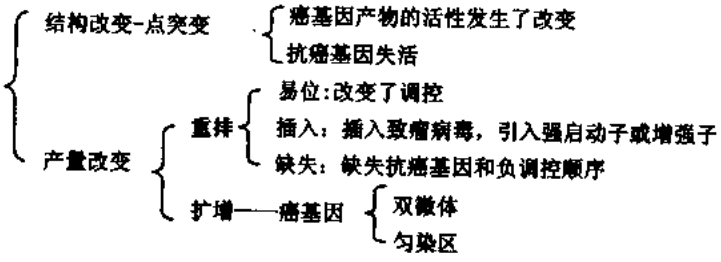


图 19－9　癌基因突变的几种情况。

与 SH2 区结合，抑制 416 位点磷酸化，受体活化后，特异位点的 Tyr 磷酸化并与 Src 的 SH2 结合，Src 的 527 从 SH2 区释放出，并去磷酸化，而 Tyr－416 位点失去抑制而磷酸化，Src 激酶活化。v-*src* 和 c-*src* 的 DNA 序列基本相同，但 v-*src* 无内含子，它比 c-*src* 少编码 7 个氨基酸(C 端)，且 C－末端编码的 12 氨基酸序列与 c－*src* 不同。这样其 527 位缺乏 Tyr 而不能磷酸化，也就不能抑制 416 位点 Tyr 位点的磷酸化，使细胞高度转化。对于 c－*src* 来说 527 位发生突变不再编码 Tyr，而是由 Phe 取代时就无法磷酸化，同样不能抑制 416 位 Tyr 的自身磷酸化，c-*src* 基因被激活，活性增加 10 倍，细胞发生转化(图 19－12)

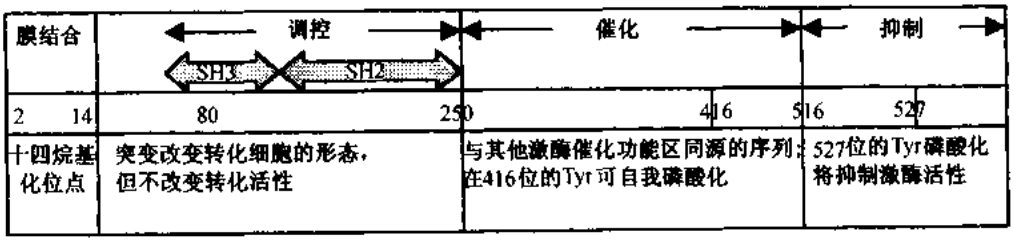


图 19－10 Src 蛋白的结构。(参考 B. Lewin: Genes Ⅵ, 1997, Fig. 37－14)

二、c-*erbB* 产物的结构与功能

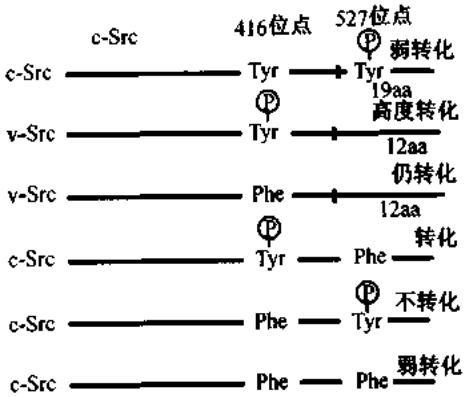


图 19－11 Src 蛋白中两类 Tyr 磷酸化位点。
(仿 B. Lewin: Genes, Ⅵ, 1997, Fig. 37－15)

c-*erbB* 编码 EGF 受体（表 19－1），是一种分子量为 65kDa 的跨膜蛋白，由 1186 个氨基酸组成。膜外的部分(621aa)是 EGF 结合功能区。前面有 24aa 构成的信号肽。跨膜区由 23 个疏水氨基酸组成。膜内部分由 542aa 组成。其中包括 4 个磷酸化位点和一个 TPK 功能区。4 个磷酸化位点中有一个是紧靠膜内侧的 654 位 Thr，其余 3 个是位于 1068，1148 和 1173Tyr 自身磷酸化位点(图 19－13)。

EGFR 是通过和配基的连接来调节的(图 19－14) 在游离的受体中膜外的 N 端功能区可抑制单体之间的相互作用，一旦和 EGF 结合就解除了这种抑制。随之受体形成了二聚体，接着 C－端的 3 个 Tyr 残基就快速自身磷酸化，激活了受体酪氨酸蛋白激酶活性，触动了信号的传递。

v-*erbB* 和 c-*erbB* 不同，它保持了 TPK 激酶功能区和跨膜区，而丢失 N 端和配基结合的功能区，C 端也丢失了第三个 Tyr 磷酸化位点，而此区域可以抑制转化活性。v-*erbB* 因丢失了配基结合区，而不受其抑制，组成型地形成二聚体，使其自身磷酸化，从而激活了 Tyr 激酶活性，这样不受配基(EGF)调节而持续地触发增殖信号，失去对转化抑制的能力，导致细胞癌变。同样原理适用于编码受体因子这一组癌基因。当它们突变时会导致错误信息的信息传递。

neu 基因是生长因子受体这一组的一员。人们从大鼠的神经母细胞病中分离得到其产物为 185kDa 的跨膜蛋白，也是 EGFR 成员，在人类中此基因定位于 17q12～21，31。它也同样有细胞外糖基化结构域，疏水跨膜区和可以自身磷酸化的 TPK 功能

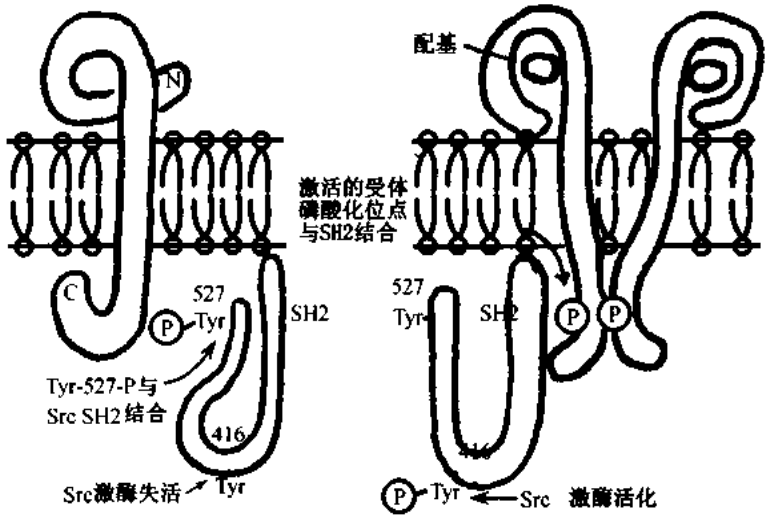


图 19－12　当受体 Tyr 激酶被激活、SH2 结合位点发生自身磷酸化，Tyr－527 被释放并去磷酸化，416 Tyr 被磷酸化，Src 激酶活化。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 37－16）

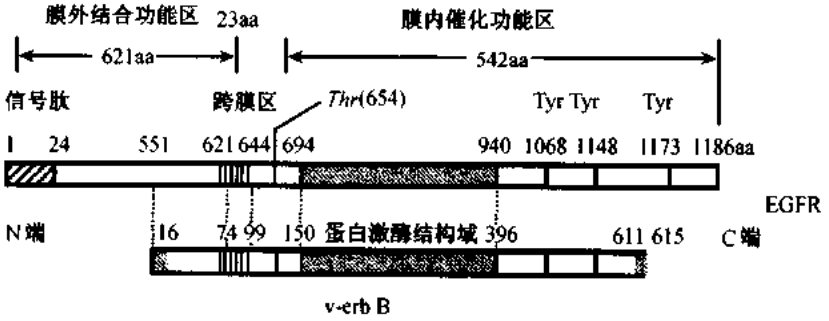


图 19－13　erb-B 癌基因产物（下）与表面生长因子受体（EGFR）（上）的结构比较。

区三个区域。在大鼠中其 664 位的氨基酸为 Val （GTG），在仓鼠中位于 658 或 659位。一旦 GTG 突变为 GAG 那么 Val 将被 Glu 所取代，其 TPK 活性增加而且导致癌变。RB 蛋白和雌激素的受体都对 *neu* 基因的转录起负调节作用。一旦它们的缺失或突变也会导致 *neu* 基因活性增强。在人类的肿瘤中 *neu* 过量表达常伴有 *neu* 的扩增或 17 号染色体的增加。

三、*c-sis* 产物的结构与功能

sis 的产物是一种分泌蛋白。1983 年 Waterfeld MD 等发现了血小板生长因子（PDGF）与猿猴肉瘤病毒的转化蛋白 p28 具有结构的相似性。Doolittle 等进一步确定 *v-sis* 癌基因来源于 PDGF 基因。它们之间有 87% （图 19－15），后来发现 c-p28c-sis 与 PDGF 序列只有 3%的差异。而人类 c-*sis* 基因产物则与 PDGF-β 链完全相同。表明病毒癌基因是来源于动物和人类细胞中的一类正常基因。

PDGF 与 p28V-*sis* 结构上的相似性表明 p28V-*sis* 可能模拟 PDGF 的作用，作为配基与膜上的受体结合，通过细胞内信号的传递，对细胞产物刺激殖作用，导致细胞的癌变。

四、*ras* 原癌基因可能因突变而被激活

直接检测癌基因有两种方法：一种是从动物肿瘤中提取 DNA 转染正常细胞使其转化，已确立的小鼠 NIH3T3 纤维瘤细胞株也可作为受体细胞。另一种方法将这种提取的 DNA 注入裸鼠（无免疫能力的一种小鼠）可形成肿瘤。用以上的方法就可用来检测 c-*ras* 家族的 *onc*。

ras 癌基因最初是在两种大鼠的反转录病毒 Ha-Musr 和 Ki-Musy（表 19－1）中发现

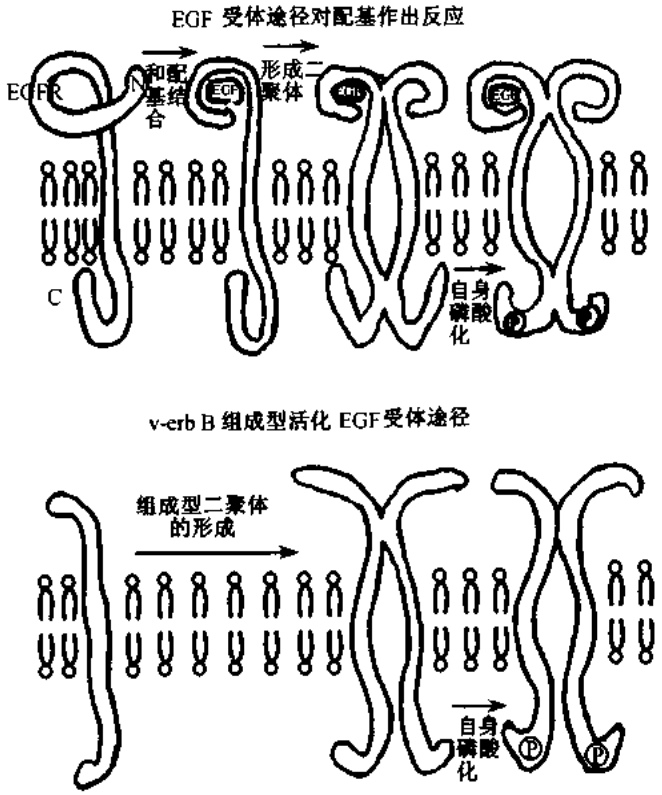


图 19－14　生长因子受体的激活涉及到配基的结合，二聚体的形成及自身磷酸化，而 V-erbB 丢失了配基结合区，由于不受 N 端功能区的抑制而组成型激活。（仿 B.Lewin:Genes Ⅵ，1997，37－13）

的。在人和大鼠的基因组中相应的 c-*ras*，N-*ras*，H-*ras* 和 K-*ras* 虽定位在不同的染色体上，但是它们互相密切相关，产物相似，分子量约为 21kDa，统称为 p21ras。V-Ras 和 C-Ras 结构相似，仅少数的几个氨基酸不同。

哺乳动物 Ras 蛋白的结构如图 19－16 所示。它有 3 个功能区负责 Ras 的特殊活性：

（1）第 5～22 残基和 109～120aa 间的区域是 GDP/GTP 的结合区，与其他的 GTP－结合蛋白同源。

（2）Ras 通过紧靠 C 端的十六烷酸附着在膜的表面上。此位点若发生阻止化学修饰的突变会被坏致瘤性，表明和膜的附着对于 Ras 的功能来说是重要的。在十六烷酸的后面有 3 个 C 端的氨基酸被剪切掉，186 位 Cys 的羧基是被甲基化的。在这附近的其

他的 Cys 残基被十六烷基化。这些特点增强了和膜的连接。

(3) 30～40 残基之间是效应区。当 Ras 被激活时此区域和靶分子起反应。这个区域对于 Ras 蛋白的致癌活性是重要的。当第 12 位点突变这个区域被激活。这个区域也是和 GAP(GTPase activation protein,GTP 酶活化蛋白)相互作用的功能区。

Ras 蛋白的结晶结构如图 19－16 所示,靠近鸟苷的区域在其他的 GTP 结合蛋白中是保守的,潜在的效应物结合区在磷酸基因的附近,由亲水氨基酸组成,它可能暴露在细胞质中。当 GTP 水解时,Ras 蛋白的构象会发生改变,此改变涉及 L4,它含有 61 位点,致癌突变有时在此处发生。Ras 的组成型激活的突变发生在 L1 的第 12 位点。它直接影响到和 GTP 的结合。在野生型和癌变型之间的变化限制在这些区域中,并妨碍突变 Ras 产生构象来承担 GTP 的水解,这是致癌性的基础,关键是 GTP 水解能力。

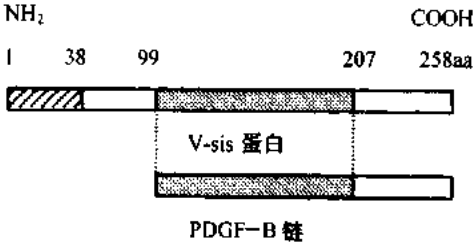


图 19－15　PDGF B 链与 v-sis 蛋白的比较。
(引自 Hunter T.,1985)

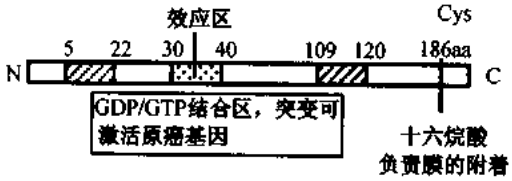


图 19－16　Ras 蛋白的结构。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 35－20)

1982 年 Tabin C J 和 Reddy E P 两组科学家在 *Nature* 上发表文章首次证实人类癌基因中的一个点突变导致了肿瘤的产生,引起了学术界极大震动。他们首先在人类膀胱癌 T24 细胞中分离出 Ha-*ras* 基因,后来又在肺癌、结肠癌中分离出 Ki-*ras* 基因,在神经母细胞瘤等细胞中获得了 N-*ras*。经测序和比较发现在各种癌组织中 *ras* 基因仅在第 12 或 61 密码上发生了点突变(病毒癌基因在第 59 位密码子上)导致氨基酸的取代(表 19－2)。

图 19－18 表明带有 GTP 的 Ras 是有活性的，可以作用其靶分子，然后内在的 GTPase 活性使 GTP 水解成 GDP，而结合 GDP 的 Ras 成为失活状态，从而终止了其作用,有两种蛋白调节 Ras 蛋白从激活状态到失活状态相互间的转变。

(1) GTP 酶激活蛋白 GAP 刺激内在的 Ras GTPase 使其水解被结合的 GTP。现已发现有多种不同的 GAPs，它们与不同的 GTP 结合蛋白特异结合。Ras－GAP 是在哺乳动物培养细胞中作用于 Ras 的 GAP(图 19－19)。

(2) SOS 是一种鸟苷释放因子(GNRF, guanine nucleotide release factor)。它刺激 GDP 被 GTP 所取代,在受体 Tyr 激酶磷酸化反应中它是被 Grb2 活化的靶蛋白。其机制可能是 Grb2 将 SOS 带到膜上，SOS 再与附近的 Ras 相互作用 (Ras 突变后内在的 GTPase 的活性减低,一直和 GTP 结合,保持了激知状态,不断作用靶蛋白,错误地将信号传递到细胞核,导细胞的癌变)(图 19－20)。

五、Jun 和 Fos 是一对转录因子

1986 年 Maki 等从鸡的肉瘤病毒 ASV17 中分离了一个新的癌基因，叫做 v-*jun*，编

表 19－2 *ras* 基因的突变和致癌性

onc	密码子			*onc* 来源
	12	59	61	
C-H-ras	GGC	GCC	CAG	正常细胞
	Gly	Ala	Gln	
H-ras	GGC	GCC	CTG	肺癌 Ha242
	Gly	Ala	Leu	
	CTC	GCC	CAG	膀胱癌 T24
	Val	Ala	Gln	
C-K-ras	GGT	GCA	CAA	正常细胞
	Gly	Ala	Gln	
K-ras	TGT	GCA	CAA	肺癌 Calu-1
	Cys	Ala	Gln	
	CGT	GCT	CAA	肺癌 LC-10
	Arg	Ala	Gln	
	GTT	GCT	CAA	结肠癌 SW480
	Val	Ala	Gln	
C-N-ras	GGT	GCT	CAA	正常细胞
	Gly	Ala	Gln	
	GGT	GCT	AAA	神经母细胞瘤 SK-N-SH
	Gly	Ala	Lys	
	GGT	GCT	CGA	肺癌 SW1271
	Gly	Ala	Arg	
V-H-ras	AGC	ACC	CAG	Ha-MuSV
	Arg	Thr	Gln	
V-K-ras	AGT	ACA	CAA	Ki-MuSV
	Ser	Thr	Gln	

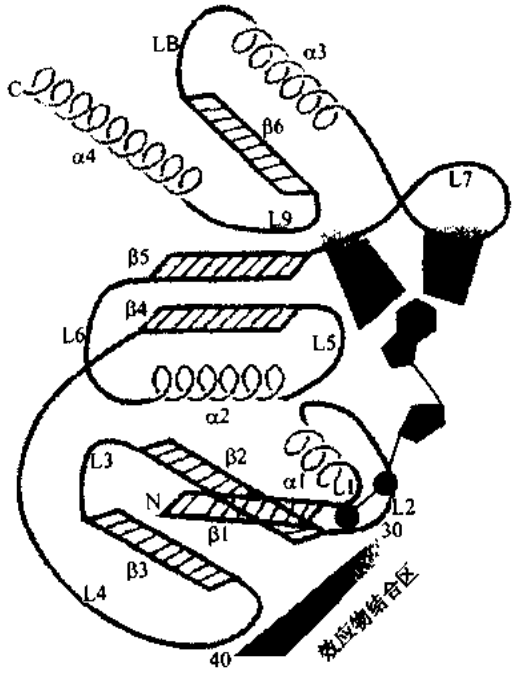


图 19－17 Ras 蛋白晶体的结构。Ras 蛋白有 6 个 β 折叠，4 个 α 螺旋，9 个连接环。L9、7、2 和 1 是 GTP 结合区。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 35－21）

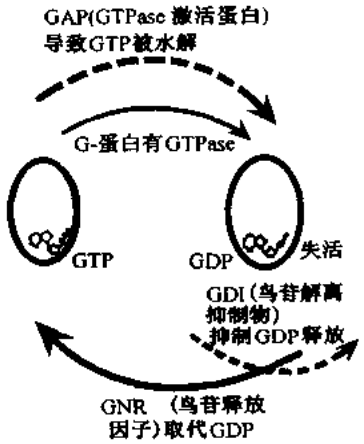


图 19－18 G 蛋白的活化及失活两种状态及相关的因子。（仿 B. Lewin: Genes Ⅵ, Fig. 35－18）

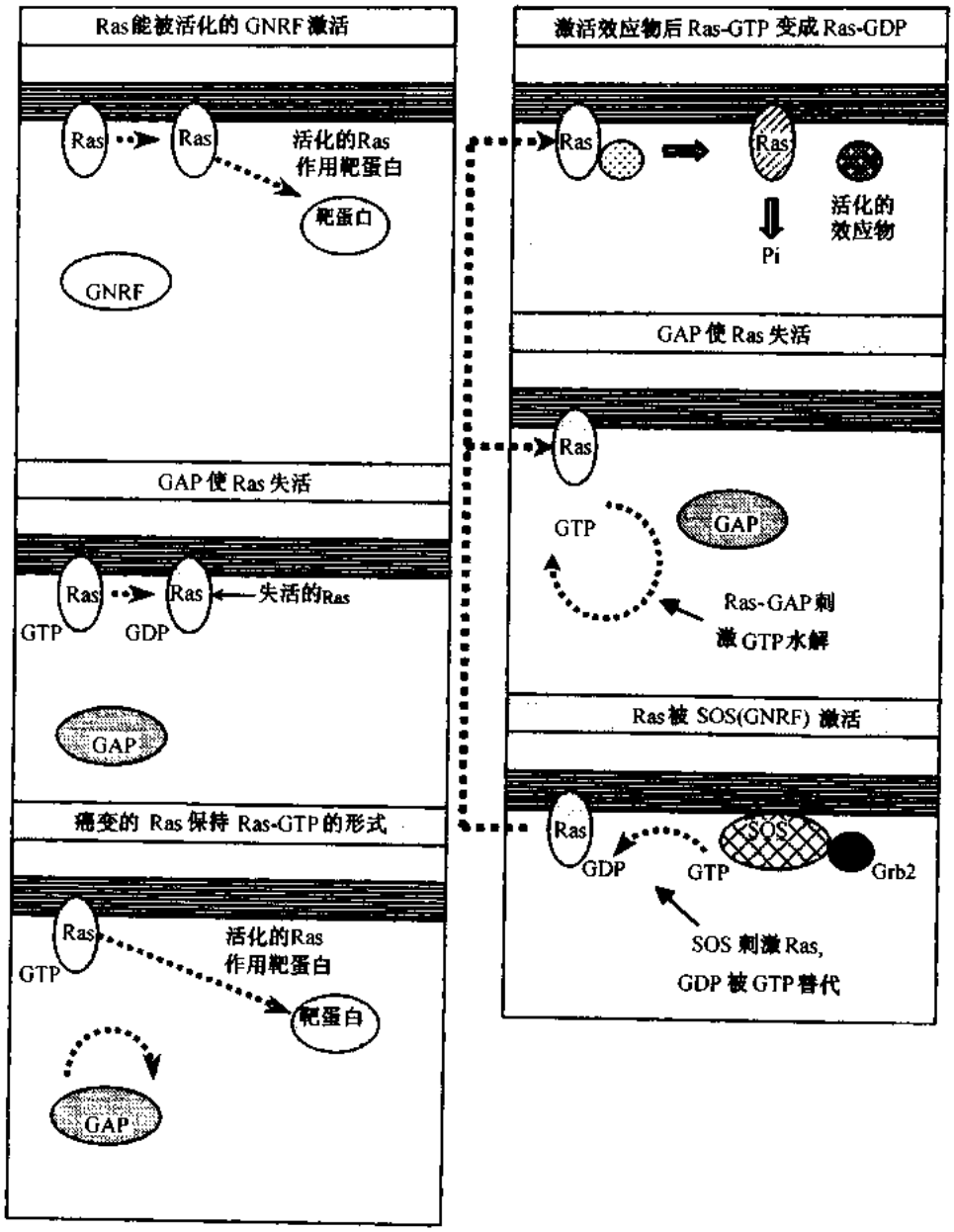


图 19－19　Ras 在正常情况下(右)和癌变后(左)的功能。GNRF:鸟嘌呤核苷释放因子;GAP:GTPase 激活蛋白;SOD = GNRF。(参考 B. Lewin: Genes Ⅵ, 1997, Fig. 35－19, 37－8)

码 296 个氨基酸的多肽,表达产物为 65kDa,是一种融合蛋白。与小鼠的 c-*jun* 有 79%的同源性,人类的 c-*jun* 的 mRNA 长 3.3kb,编码 39kDa 的核内磷蛋白 $p39^{c\text{-}jun}$,人类的 c-*jun* 和小鼠的 c-*jun* 有 90%的同源性。

人类的 c-*jun* 基因定位于 14q21～31,成熟的 mRNA 长 2.2kb,编码含 380 个氨基酸的 $p55^{c\text{-}fos}$,分子量 55kDa,也是一种核内磷蛋白。

AP－1 是从 HeLa 细胞中提取的多肽,分子量为 39kDa,是一种转录因子,识别增强子和某些启动子上的基序 TRE[TPA 佛波酯(phorbol ester)]应答元件(TPA response element),保守序列为 TGACTCA。1988 年 Angel 等证明 *jun* 癌基因的产物与 AP－1 有一定的相似性;Franza 等报告 Fos 蛋白的复合物与 DNA 结合的保守序列也和 AP－1 一致。

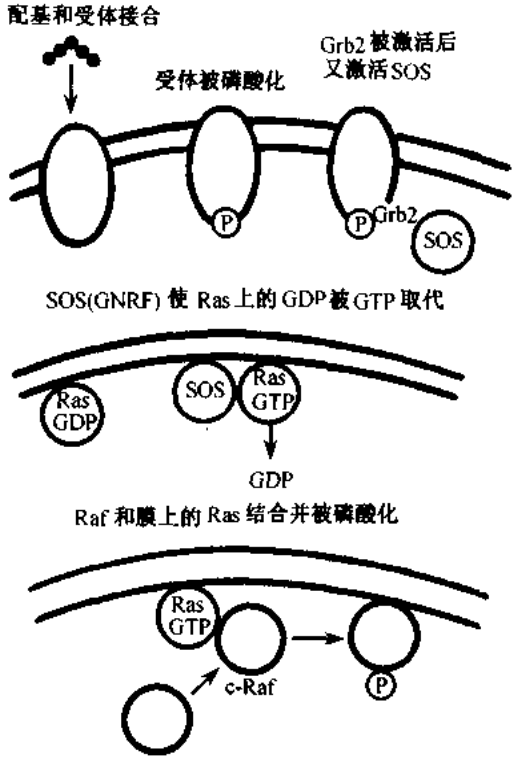


图 19－20　Ras 的级联反应是从一系列激活反应开始的它发生在质膜内侧。(仿 B. Lewin: Genes Ⅵ, Fig. 35－17)

现已搞清 AP－1 因子由二个亚基组成异二聚体，它们正是 c-*jun* 和 c-*fos* 编码的产物。以亮氨酸拉链的形式调节具有 TRE 的靶基因(图 19－21)。

现在还不清楚 c-*fos* 和 c-*jun* 发生什么样的变化会导致细胞的癌变。推测可能涉及二者产物数量的改变，以致使靶基因表达过量或减少，使细胞发生转化。

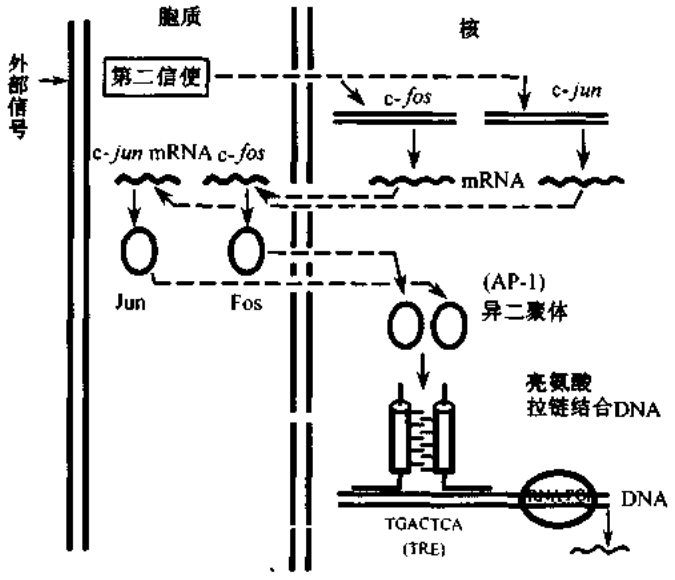


图 19－21　c-*fos* 和 c-*jun* 的功能。

第四节　染色体畸变和肿瘤的发生

细胞癌基因都具有正常的生理功能。维系细胞中信息的顺利传递。基因突变会改变其产物的结构或表达调控模式而导致肿瘤的发生。染色体畸变也会由于染色体的重排和基因的扩增而使细胞癌变。*Rb*(视网膜母细胞瘤，retinoblastoma) 基因和 *P53* 抗癌基因是缺失导致肿瘤产生最有名的例子，而 c-*myc* 基因表现了癌基因的 3 种激活途径:反转录病毒的插入,染色体的易位和基因的扩增。三者的共同特点是癌基因表达的增强,而不是其结构的改变。虽然在少数情况下其转录本丢失了不翻译的前导序列,也并不影响其产物结构。c-*myc* 提供了一个通过增强或改变表达而被激活的癌基因的例子。

一、*RB* 和 *P53* 的缺失导致肿瘤的发生

人类的视网膜母细胞瘤 (Rb) 原被认为是常染色体显性遗传的恶性肿瘤、患者常见于幼儿，发病率为 1/（1.7～3.4 万)。约有 5% Rb 患者检测到 13 号染色体长臂缺失 ($13q14^-$),现在已弄清在 13q14 有一抗癌基因 *RB-1*,编码 928 aa 的核内磷蛋白,可抑制细胞增殖,当其缺失或突变不仅会引起 *RB* 的产生,还会导致骨癌和小细胞肺癌的产生。

当一对 *RB* 基因都失活才会导致视网膜母细胞瘤，在遗传型的病例中，个体常常是缺失了一个 *RB*,还带有一个正常的 *RB*。但如果其视网膜细胞作为体细胞丢失了另一个 *RB* 基因拷贝的话就会产生肿瘤。在散发的病例中亲代的染色体是正常的，但患者的体细胞中丢失了一对 *RB* 等位基因。

几乎半数的 *RB* 患者都是缺失了 *RB* 座位。另一些病列虽具有 *R*b 座位,但不是不能转录,就是发生了突变。在 *RB* 细胞中没有 Rb 蛋白产物的,或者其产物失去了正常功能。

除了 Rb 以外还有其他一些肿瘤其发病也和染色体缺失有关，如神经母细胞瘤 (neuroblastoma)，黑色素瘤 (melanoma)，肺小细胞癌 (lung, small cell Ca)，威尔姆氏 (Wilm's) 瘤(一种肾脏恶性胚胎瘤)，睾丸瘤(testicular tumors)等一些人类的实体瘤患者都发现部分患者存在相应染色体的缺失(图 19－22)。

二、反转录病毒的插入激活 c-*myc* 癌基因

在人类中，病毒的插入也会引起某些肿瘤的发生。如 c-*myc* 原癌基因编码一种转录因子,行使基因调控的功能。但一种缺陷型强致癌的反转录病毒 ALV 感染细胞插入到 c-*myc* 座位后就可以不同的方式激活这个癌基因,产生游离的肿瘤。c-*myc* 基因由 3 个外显子组成，最前面的一个长的不翻译的前导序列，后面是二个 c-*myc* 蛋白的编码区。图 19－23 表明 c-*myc* 插入序列的类型。反转录病毒可在不同位置插入 c-*myc* 基因。

最简单的情况是 ALV 插入在第一个内合子中。LTR 提供了一个启动子,以通读的形式转录了 2 个编码的外显子。在 ALV 的控制下 c-*myc* 的转录与正常情况不同:(1)转录水平超过了正常转录，这是由于 LTR 的启动子是一种高效启动子。(2) 转录本中没

有正常翻译的前导序列。

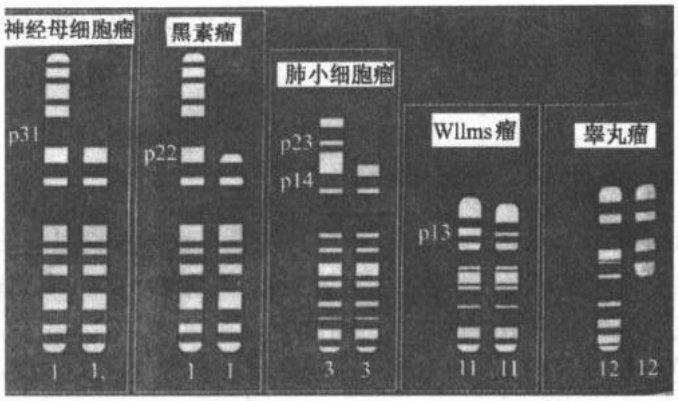


图 19-22 在人类几种不同类型实体瘤中常发现存在染色体缺失。带的数目表明存在断裂点。(引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 8-12)

插入序列对 c-*myc* 的另两种活化形式反映了不同的机制。反转录基因组可以插入第一个内含子内，但方向可以不同。其中有一种排列其启动子的方向和 c-*myc* 基因的方向相反。有可能病毒的长末端重复序列(LTR)提供了增强子，可以作用于其上游的

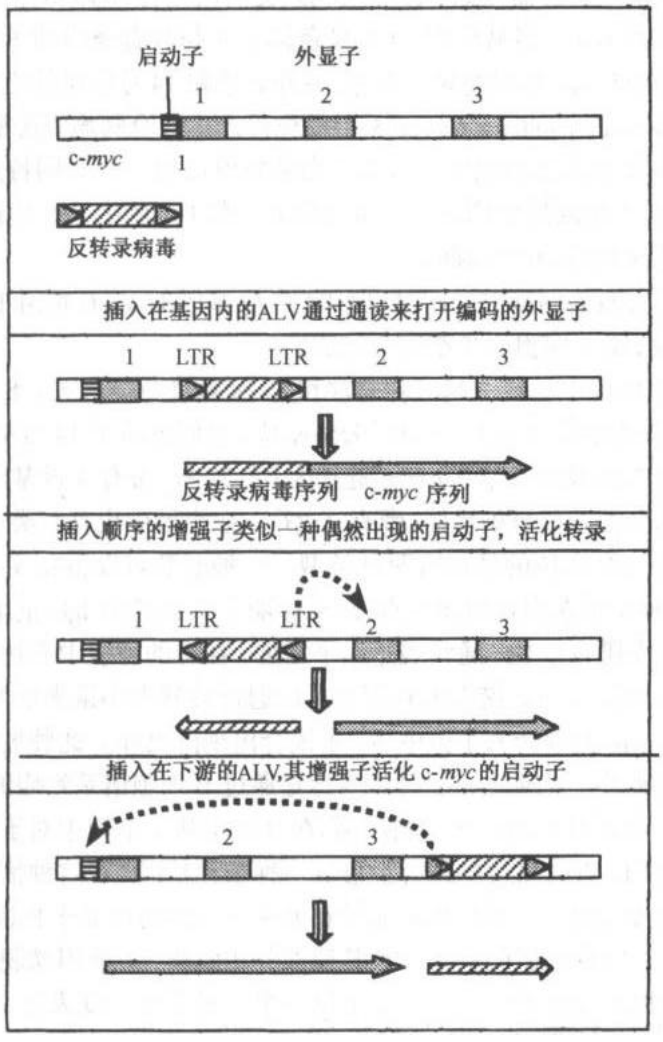


图 19-23 插入在 c-*myc* 基因不同位置上的 ALV，以不同方式活化基因。(仿 B. Lewin: Genes VI, 1997, Fig. 37-9)

序列，相当于偶然出现的启动子。反转录病毒的基因组也可插入在 c-*myc* 的下游，但其 LTR 的增强子对 c-*myc* 的启动子作用，增强了转录。

在以上的各种情况下 c-*myc* 的编码序列未发生改变，致癌性是由于失去了正常的控制，增强表达所致。

三、易位

1960 年 Nowell 和 Hungefora 在美国费城发现慢性粒细胞白血病的患者中有一个很小的近端着丝粒染色体，人们就称此为费城染色体(Philadelphia chromosome，Ph)，Ph 染色体已被公认为慢性粒细胞性白血病的特异标记染色体。Riodan O(1971)用荧光带法鉴别出 Ph 染色体是 22 号染色体长臂丢失了 1/3 而形成。Rowly(1973)进一步发现缺失的染色体片段易位到 9 号染色体上。

1982 年 De Klein 等又在 Ph 染色体上发现有 9 号染色体的片段，上面带有癌基因(c－*abl*)，表明 Ph 是相互易位的产物，癌基因 c-*abl* 可能由于位置的改变而导致表达异常，从而使细胞产生了癌变。

1971 年胡德发现小鼠 B 淋巴细胞 15 号染色体含有一个癌基因，当易位到 12 号染色体时大量合成畸形蛋白而致癌。1972 发现人类的白基特(Buvkitt)淋巴瘤细胞中 8 号染色体有原癌基因 *myc*，当易位到 14 号染色体上时与含有强启动子的免疫球蛋白重链基因重排，可促使 *myc* 基因的异常表达。另外 6 号和 14 号染色体之间的易位会产生卵巢乳突状癌(ovarian papillary Ca)，t(3;8)易位可产生混合腮瘤等肿瘤(图 19－24)。

上面谈到小鼠 B 淋巴细胞癌和人类的白基特淋巴瘤，其共同特点是在 B 细胞中一条染色体上的 *Ig* 基因易位到另一条染色体上。在 T 细胞中 *TcR* 座位(T 细胞受体，T cell recepter)发生易位也会产生肿瘤。

易位的结果在激活癌基因的同时也影响了 *Ig* 基因的 V、C 重组（见第 23 章），或 IgHC 的类别转换，使 B 细胞或 T 细胞不能成熟。

在人类 B、T 细胞中易位引起的肿瘤常和 8，14 号染色体有关。8 号染色体上带有 c-*myc*，而 14 号染色体带有 IgH 座位和 *TcRα* 座位，它们分居于 14 号染色体的两侧（图 19－25），约有 10%涉及到 2 号(带有 κ 链基因)或 22 号(带有 λ 链基因)染色体。在 T 细胞中的易位还会涉及 7 号染色体(带有 *TcRβ*)。在小鼠中也存在类似的情况。

在 B 细胞中 *IgH* 座位的易位有两种类型，一种断裂点发生在 V－D－J 重组的保守序列中，另一种发生在类别转换位点，结果影响了 B 细胞的 IgH 重组和类别转换。

当 c-*myc* 易位到 Ig 座位，通常表达水平增加，在有的肿瘤中表达水平的变化范围很大，约为 2～10 倍。c-*myc* 潜在的致癌性可以通过转基因小鼠来证明，将与 IgH 的增强子相连锁的 c-*myc* 基因转入小鼠中后，小鼠会出现淋巴瘤。此肿瘤存在于成熟和未成熟的 B 淋巴细胞中。表明 c-*myc* 过量表达是通过 B 细胞谱系致瘤的。

癌基因常在交互易位的产物中被激活。在 B 细胞和 T 细胞中对于各种易位已鉴别出一些新的癌基因。有时易位的结合会形成一种杂种基因。在这种情况下原来的转录单位被易位所改变，使一个基因的外显子和另一个基因的外显子相接，这就存在两种潜在的致癌性：①原癌基因的一部分和其他部分相脱离，可能因摆脱某种控制而被激活。例如在新的控制条件下：c-*myc* 失去了第一个外显子而过度表达。②易位后与癌基因融合的另一个基因的部分序列可能有正调节功能，如增强子，结果增强了表达。

前面提到的 Ph 染色体就是一种交互易位产生杂种癌基因的例子。其一部分来自

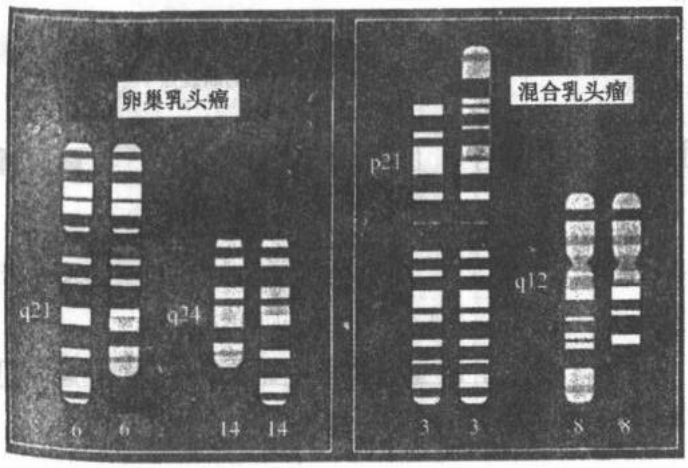


图 19-24 在人类几种不同类型实体瘤中常发现存在染色体易位的现象。带的数目能表明断裂点。
(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 8-31)

9号染色体片段，长 5000kb，它的末端带有 c-*abl* 区域，它与 22 号染色体上的 *bcr*(断裂点簇区域，break-point claster region)相接，*bcr* 以前描叙为 22 号染色体上常发生断裂的一个长约 5.8kb 区域。在 CML 中此区域的易位连接点有两种（图 19-26)，易位的结果 *bcr* 区域位于一个大的融合基因中(790kb)，现在称 *bcr* 基因。在 CML 中断裂点常发生在基因中部两个内含子的一个中。这个基因也因易位产生另一种血癌叫 ALL(急性淋巴白血病，acute lymphoblastic leukemia)，在此病中断裂点发生在 *bcr* 基因的第一个内合子中。c-*abl* 与 *bcr* 有不同的选择性剪接。易位产生了编码 *Bcr-Abl* 融合蛋白的转录本。其 N 端序列是 *bcr*，C 端是 c-*abl*。在 ALL 中 185kDa 的融合蛋白中有 45kDa 的 Bcr 蛋白和 140kDa 的 c-Abl 蛋白，(c-Abl 蛋白完整的部分为 154kDa)，也就是说它丢失了少量(4kDa)的 N 端的氨基酸。

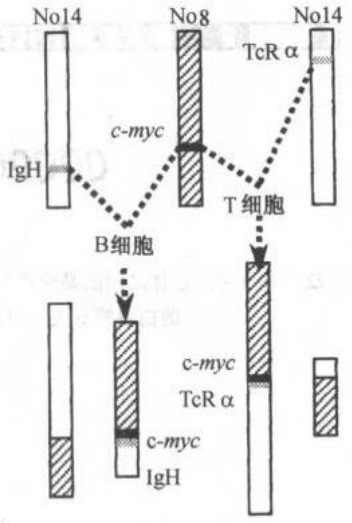


图 19-25 易位激活了原癌基因。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 37-10)

为什么这种融合蛋白会致癌呢？那是由于 N 端的 Bcr 区域和 c-Abl 蛋白相互作用。Bcr 有一个可变性的基序，它和信号传递有关，具有 Ser/Thr 激酶活性。在此部位的蛋白有自我磷酸化残基。磷酸化后可与 c-Abl 的 SH2(肉瘤基因同源区 2，*src* homology region-2)功能区相互作用而改变构象，从而激活了 c-*Abl* 潜在的致癌性(图 19-26)。

N 端结构的改变也涉及到 v-*abl* 致癌性的激活，v-*abl* 是反转录病毒所携带的癌基因。而相应的 c-*abl* 基因编码的是一种 Tyr 激酶蛋白。N 端实际上是一个调节激酶活性的区域，一旦丢失将导致癌基因的激活。

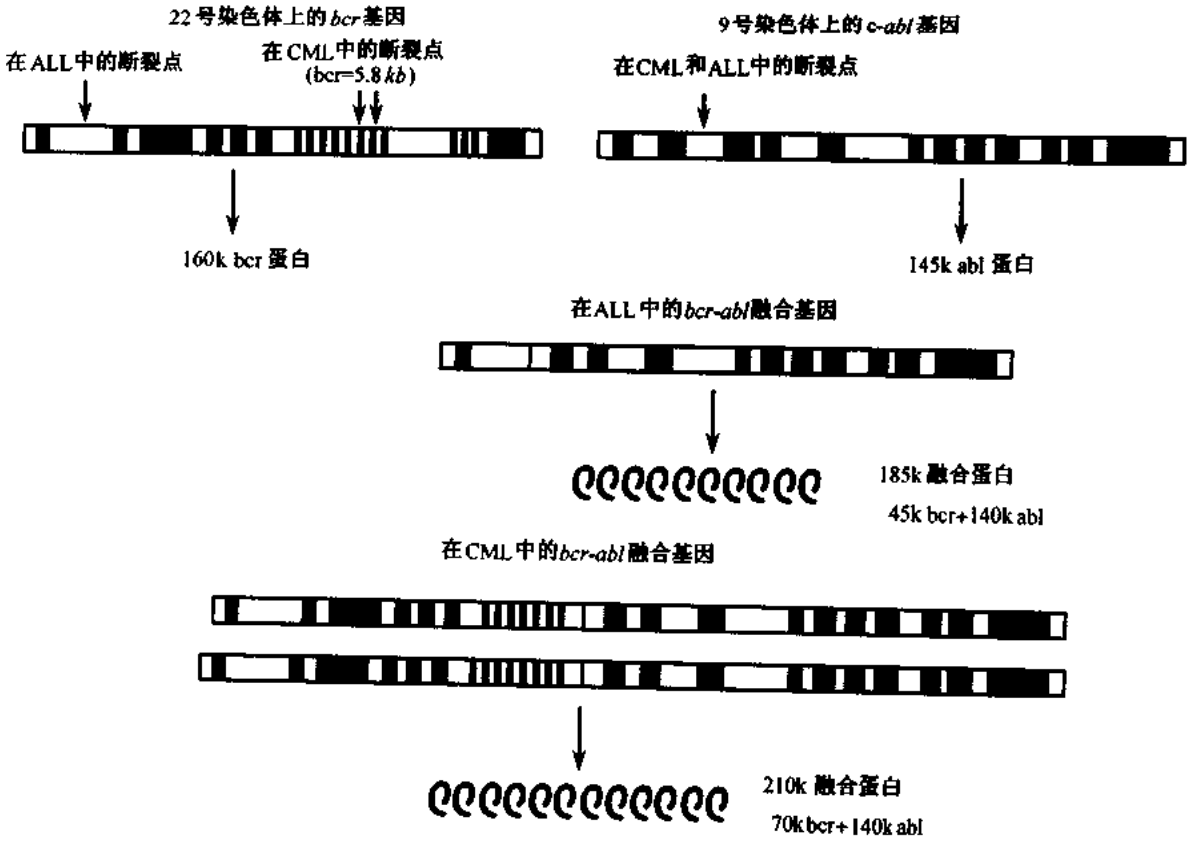


图 19－26　在 22 号和 9 号染色体之间的易位产生了费城染色体，它合成 bcr－abl 融合转录本。不同的融合基因和不同的白血病有关。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 37－11）

第二十章

核外遗传信息的传递

基本原理

线粒体和叶绿体的基因组都是双链环状的 DNA。在两种细胞器中含有多个类核区，每个类核区含有多个拷贝的 DNA 分子。

线粒体和叶绿体的基因组都含有作为核糖体成分 rRNA 的基因。线粒体和叶绿体中的核糖体是在细胞器中装配和行使功能，在细胞器中有很多 tRNA 参与合成细胞器蛋白。细胞器中的蛋白少部分是在细胞器中合成的，它们往往具有特殊的功能。其他大部分蛋白都是由核基因编码的，在细胞质中的核糖体上合成，然后再转运到细胞器中，这些蛋白包括细胞器核糖体的蛋白。

已经知道很多线粒体和叶绿体突变的例子，这些核外基因的遗传的规律不同于核基因。不涉及减数分裂分离，常出现单亲遗传（常是母本），核取代后核外突变产生的性状仍存在，核外基因不能像核连锁群那样作图。

母体影响定义为在卵受精前其核基因控制的性状由母体的核基因型提前决定。母体影响不同于核外遗传。核外基因的母体遗传的模式是由于受精卵的大部分细胞器是从母本那儿获得的。相比之下，母体影响的遗传性状是在卵受精前由母体的核基因组控制的，不涉及核外基因。

有很多核外遗传基因突变的例子涉及到线粒体和叶绿体的 DNA，包括洗澡花（紫茉莉）的花斑枝条的遗传，真菌小菌落突变以及衣藻对抗菌素的抗性和敏感性等。

不是所有的核外遗传都是由线粒体 DNA 和叶绿 DNA 编码的。在真核中有很多的例子是由传染遗传所致，当细胞质混合时位于细胞质中的病毒或细菌会随之转移。

前面我们已分析了染色体上核基因的结构及表达，并了解其遗传的规律，但 DNA 并不仅位于核中，在细胞质中也有少部分 DNA 的存在，主要位于动植物的线粒体和绿色植物的叶绿体中。线粒体和叶绿体的基因组称染色体外基因，或细胞质基因，非孟德尔基因，细胞器基因及核外基因等。称其为非孟德尔基因就表明核外基因的遗传并不像核基因那样的分离与重组。近年来现代分子生物学技术的发展和应用已搞清了核外基因组的结构。

第一节　核外基因组的结构

一、线粒体基因组

线粒体是好氧动物和植物细胞质中的细胞器，是细胞中能量的来源，它们含有三磷酸循环的酶类，执行氧化磷酸化及脂肪酸的生物合成。通过电镜、放射自显影、浮力密度比较等方法对线粒体的 DNA 进行了系统的研究，线粒体 DNA（**mtDNA**）的一些特点是如表 20－1 所示。

不同物种的真核生物线粒体基因组大小差异很大，但含量是恒定的，高等植物 mtDNA 链长 250kb～2000kb（80～800μm）。玉米的 mtDNA 基因组有 600 kb（200μm）。真菌、昆虫直到哺乳动物显示出一种趋势，即从低等到高等直核生物 mtDNA 越来越小，排列越来越紧密。如酵母：约 75kb，粗糙脉孢菌：60kb；曲霉：32kb；爪蟾：17533bp；牛：16338bp；小鼠：16295bp；人类仅 16569bp。链长的差异是由间隔区的内合子的数目和长度决定的，以上表明高等生物 mtDNA 基因排列是最经济的，这和核基因组的情况正好

表 20－1　核 DNA，mtDNA 和 cpDNA 的性质

		核 DNA		mtDNA		cpDNA	
		% GC	基因组大小(bp)	% GC	基因组大小(bp)	% GC	基因组大小(bp)
动物	人	41	2.75×10^9	41	14×10^3	–	–
	鸡	43	1.2×10^9	49	15×10^3	–	–
	海胆	39	7.2×10^8	45	13×10^3	–	–
	果蝇	42	1.75×10^8	22	18×10^3	–	–
真菌	脉孢菌	53	4.3×10^7	42	60×10^3	–	–
	酿酒酵母	44	1.75×10^7	18	75×10^3	–	–
植物	*C. reinhardi*	64	1×10^8	71	2×10^5	36	30×10^3
	玉米	49	6.6×10^9	48	2.2×10^5	39	13×10^3
	烟草	38	1.1×10^9	51	2×10^5	47	15×10^3
	香豌豆	35	9.4×10^9	38	9×10^4	46	12×10^3

Peter J. Russell: Genetics, 3rd ed., 1992, Table 21.1

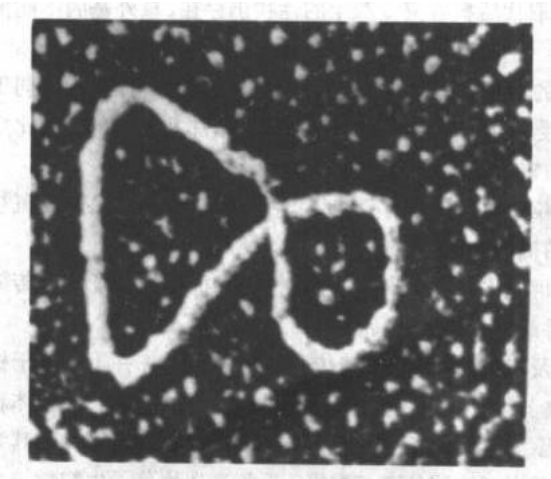


图 20－1　线粒体 DNA 的电镜照片。
(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 21－2)

相反。

mtDNA 是双链环状的分子，由于缺乏组蛋白，故不形成核小体(图 20－1)。在线粒体中有和细菌细胞中相似的类核区(nucleocid regions)，每个类核中含有几个拷贝的线粒体染色体。例如酵母每个线粒体含有 10～30 个类核，每个类核含有 4～5mtDNA 分子。由于每个酵母细胞中有 1～45 个线粒体，这样每个细胞就有很多 mtDNA 分子。线粒体染色体的多重性对于细胞的氧化磷酸化活性显得更为重要。

1. mtDNA 的基因组成

线粒体 DNA 含有大量线粒体成分的遗传信息，如 tRNAs，rRNAs 和蛋白。线粒体含有的核糖体是负责线粒体中所有蛋白质合成的。mRNA 也是在线粒体中合成的，它保留在细胞器中，由线粒体的核糖体进行翻译。在线粒体的核糖体中有两类 rRNA 分子，它们由 mtDNA 编码。线粒体的核糖体蛋白大部分由核基因编码。从测得的几种脊椎动物 mtDNA 全序列来看，它们结构的共同特点是：①基因数目和排列顺序相同。共有 2 个 rRNA 基因(12S 和 16S rRNA)，22 个 tRNAs，13 个线粒体膜蛋白基因和 1 个氧化代谢所需的基因；②其中 2 个 rRNAS，14 个 tRNAs 和 12 个蛋白质基因在重链上(H)，另 8 个 tRNAs 和 1 个蛋白质基因在轻链(L)上(表 20－2)；③有一个 D 环，与 mtDNA 的复制有关；④有 2 个复制起始点，分别复制 H 和 L 链；⑤基因间没有间隔，因此每个基因不可能都有自己的启动子，明显具有原核生物性质；⑥某些蛋白质的密码子与核基因通用密码子不同(表 20－2)；⑦mtDNA 主要编码 rRNA 和 tRNA 分子，氧化呼吸所需要的酶类大部分亚基由核基因编码，mtDNA 仅编码其中少部分亚基。

蛋白质基因位于两条链上是怎样鉴别的呢？人们通过两种方法，一种方法是分析同一个开放阅读框的起始和终止信号序列，它们中间必定是编码区，人们用计算机进行分析，解决了这个问题；第二种方法是由 G.，Attardi 小组采用的，他们将线粒线的 mRNA 和 DNA 相对比来鉴别其 5′端和 3′端（线粒体的 mRNA 5′ 无帽结构，但 3′端有 poly（A）尾巴）。

1980 年人类 mtDNA 中所有的未知的开放阅读框（unidentified reading frame，URF）都被成功地鉴别出来了。人们小心地分析基因序列，揭示出 mtDNA 中的密码子与核基因密码子的某些差异。①AUA 成为 Met 的密码子，而不是通用 Ile 密码子；②UGA 是 Trp 密码子，而不是终止密码子；③AGA，AGG 在哺乳动物 mtDNA 中是终止密码子，而不是 Arg 密码子。在酵母的 mtDNA 中这些密码子的编码又有所不同（表 20－2）。此表明 mtDNA 遗传密码在许多真核生物中可能有不同的阅读方式。另外遗传密码并不像原来想的那样通用。

用计算机来鉴别人类 mtDNA 的 tRNA 基因，其 tRNA 为典型的三叶草二维结构，其特点和细胞质中的 tRNAs 是相同的。在细胞质中密码子和 tRNA 的反密码子相互匹配时存在着碱基的摆动（见第 14 章）。61 个有义密码子最低只需要 32 种 tRNA 就可以全部阅读了。但没有一个细胞质 tRNA 能阅读 3 上以上的相关密码子。所以同一个氨基酸的 4 个密码子的前两个碱基是相同的，它们组成了**四密码子家族盒**（four-codon family

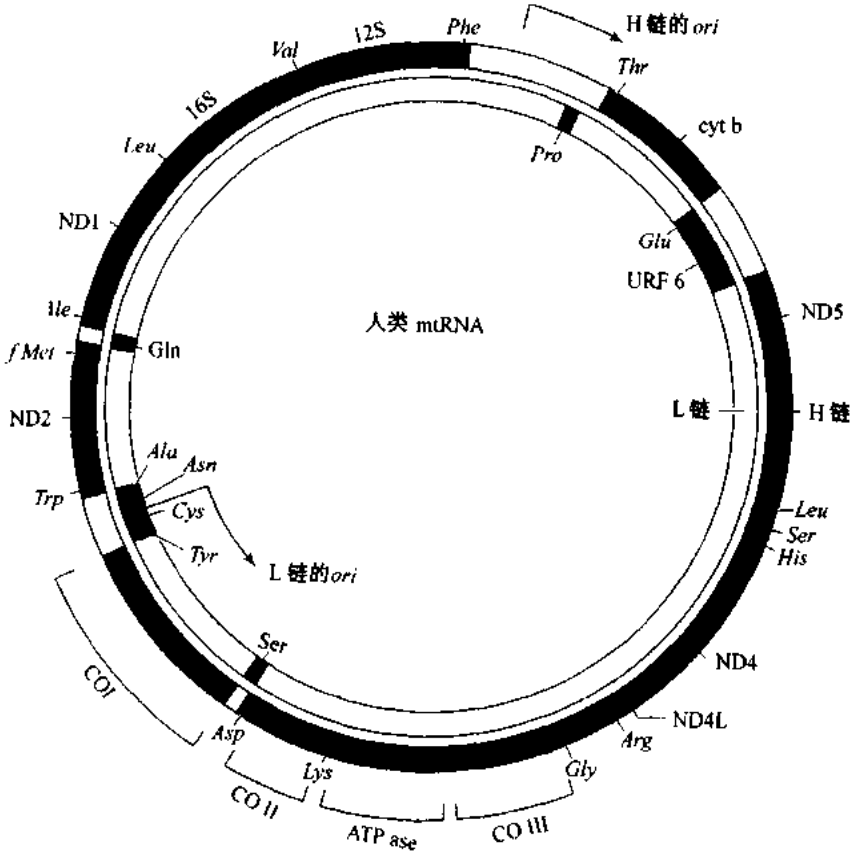


图 20－2　人类线粒体 DNA 的基因图。外圈表示从重链转录；内圈表示从轻链转录。O_H 是重链复制起始点；O_L 是轻链复制起始点。16S 和 12S 是 rRNA 基因。小格子表示 tRNA 基因，其余的框表示蛋白编码基因。它们编码 ATPase 复合体。CO Ⅰ，CO Ⅱ，CO Ⅲ：细胞色素 c 氧化酶亚基；cyt：表示细胞色素 b。ND1－6：NADH 脱氢酶成分。（引自 Russell，P. J.：Genetics，3rd ed.，1992，Fig. 21－6）

表 20-2　在线粒体中密码子的改变

生物	密码子	mtDNA 编码	核密码
共同	UGA	Trp	终止
哺乳动物	AGA/AGG	终止	Arg
哺乳动物	AUA	Met(起始)	IleIle
果蝇	AUA	Met	Ile
果蝇	AGA	Ser	Arg
纤毛虫	UAA/UAG	Glu	终止
另一类纤毛虫	UGA/UAG	Cys	终止
酵母	AUA	Met(延伸)	Ile
酵母	CCA	Thr	Leu
酵母	CUG	Ser	Leu
酵母	UAC	Ser	终止
支原体	UGA	Trp	终止

表 20-3　Ala 的 tRNA 反密码子摆动

密码子	反密码子	
	核 tRNA	线粒体中的 tRNA
GCU	GGC	UGC
GCC		
GCA	UGC	
GCG		

Peter J. Russell: Genetcs, 3rd ed., 1992, Table 21-3

box)，由两个 tRNAs 来阅读。而很多的线粒体 tRNAs，可以阅读四密码子家族盒中所有的 4 个密码子（表 20-3）。根据这种情况的推测，在哺乳动物线粒体中只需要 22 个 tRNA 基因。在研究粗糙脉胞菌和酵母的线粒体时也发现相同的情况。

前面已叙述每个线粒体基因组含有二个 rRNA 基因，但在不同的生物中这两个基因组成不同。如在人类的 HeLa 细胞中，小的（12S）和大的（16S）rRNAs 紧密地连接在一起，在其间隔中有一个 tRNA 基因。而在链孢霉中两个基因相隔 6kb，在酵母中约相隔约 30kb。大量的 tRNA 位于间隔区。链孢霉和酵母的 rRNA 基因中含有内含子（图 20-3）。在原生和动物四膜虫（*Tetrahymena*）中有一个额外的大 rRNA 基因拷贝，但小 rRNA 基因却没有。这一额外的 rRNA 基因是由于重复产生，

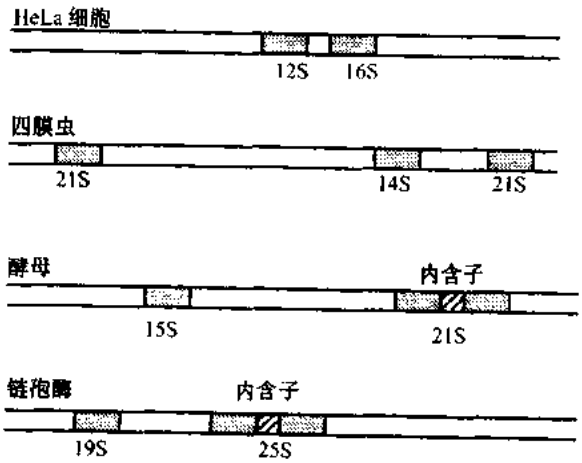


图 20-3　在 HeLa 等真核细胞中 mtRNA 的组成。
（仿 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 21-8）

这在线粒体基因组中并不多见。

人类 mtDNA 的一个显著特点是有效地组成了一个紧密结构。蛋白编码基因和 rRNA 基因连接，其间很少或无间隔存在。线粒体表达产生一个很大的转录本。通过加工，tRNA 转录本序列被剪下，它们的三叶草结构是很容易被识别的。在大 RNA 前体分子中 mRNA 和 tRNA 的转录本之间没有间隔，这一过程产生了成熟 tRNA 和单基因的 mRNAS。

在线粒体基因组中，mRNA 的起始密码子 AUG 在分子的 5′端，而细胞质中的 mRNA 是位于 5′端前导序列之后。线粒中 mRNAs 也没有 5′帽结构。胞质中的 mRNA 对于翻译来说必须要有 5′帽子和位于内部的起始密码子 AUG。线粒体 mRNA 的结构特点表明线粒体蛋白合成装置与细胞质中的核糖体是不同的。其起始机制可能是简单地识别 mRNA 的末端，而并不需要特殊的核糖体结合位点或 5′帽子结构。mtRNA 的翻译起始机制也不同于细菌，它无需要位于起始密码子 AUG 上游的核糖体结合位点——S－D 序列。

人类 mtDNA 高效组成的另一特点是 mtDNA 对于大部分线粒体 mRNA 都没有编码链终止的密码，取而代之的是在转录本的末端带有 U 或 UA，而在 mRNA3′端加 poly(A)尾巴补充 UAA 终止密码的丢失部分。虽然这种有效组成在动物 mtRNA 中占优势，但在带有较大的线粒体基因组的生物中情况完全不同。在酵母 mtRNA 中的基因组，要比编码同样成分的序列大 5 倍多，这些基因被长的无生物功能的 AT 丰富序列所分开。在这些生物的某些 mtDNA 基因中也发现了内含子，但在动物的线粒体基因组中从未发现过内合子。

2. 线粒体核糖体

生物中线粒体核糖体的结构不同于胞质核糖体，它们的蛋白和 RNA 的比例要比 *E.coli* 核糖体的高得多，表 20－4 表明人类和酵母线粒体核糖体的结构特点。在动物中线粒体糖体大小大约相同，其分子量和体积略大于 *E.coli* 的核糖体。由于高蛋白/RNA 比值，所以浮力密度相对较低，约为 55S～60S。在真菌中(如酵母和链孢霉中)线粒体核糖体的沉降系数为 70S～75S。

在大部分情况下，线粒体核糖体的 tRNA 缺乏 5S 和 5.8S rRNA。高等植物的线粒体核糖体有 5S RNA 分子，但和细胞质中的不同。大部分 mt 核糖体都有两个大小不等的亚基。每个亚基带有单个的 rRNA 分子。在动物的线粒体中这些 rRNA 分子的沉降系

表 20－4 人类和酵母线粒体核糖体和细胞质核糖体的大小

成分	人类(HeLa)		酵母	
	mt	cyto	mt	Cyto
核糖体	60S	74S	75S	80S
大亚基	45S	60S	53S	60S
小亚基	35S	40S	35S	40S
核糖体 RNA				
大亚基	16S	28S	21S	26S
小亚基	12S	18S	15S	18S

mt：线粒体，cyto：细胞质。Peter J. Russell: Genetics, 3rd ed., 1992, Table 21.4

数分别为 16S 和 12S，明显比 *E.coli* 两个亚基中主要 rRNA(23S，16S)要小。

线粒体核糖体蛋白的数目未能确定。酵母核糖体中有 60～70 种蛋白，动物核糖体的蛋白约在 70～80 种之间。

线粒体核糖体蛋白不同于细胞质核糖体的蛋白，可能除链孢霉和酵母以外线粒体蛋白都是由核基因编码，在细胞质核糖体上合成，然后转运到线粒体中。这一过程再次表明细胞能将蛋白合成的产物导向所需要的特殊地方，我们知道真核细胞核糖体含有约 65～70 种核糖体蛋白，而在低等真核生物的线粒体核糖体中也有 60～70 种核糖体蛋白。那么在细胞质中要产生 125～140 种核糖体蛋白，其中约有半数要迁到细胞核中，用于胞质核糖体的装配核糖体的装配，余下的就运入线粒体，用于线粒体核糖体的装配以及和位于线粒体染色体的两种 rRNA 基因转录本结合。

3. 线粒体蛋白质的合成

线粒体有它们自己特有的蛋白合成装置，以细菌合成蛋白质的相似方式生产蛋白质。在线粒体中存在特异的 fMet－tRNA 用于蛋白质合成起始。起始因子(IFs)也和细菌中的十分相似。在很多的情况下线粒体蛋白质合成体系的成分和细菌的相关成分在体外可以互换。例如链孢霉的核糖体能识别、结合和翻译 *E.coli* 的 AUG，而且 *E.coli* 的 IFs 可催化蛋白合成的起始。

E.coli 和线粒体核糖体之间相似性的另一条证据是后者对细菌核糖体功能的大部分抑制剂都是敏感的，其中包括链霉素、新霉素和氯霉素等。在有的情况下对这些试剂(大部分是抗生素)的敏感程度与 *E.coli* 不同。线粒体核糖体对有些抗菌素不敏感，但细胞质核糖体对这些抗菌素是敏感的，如放线菌酮。通过选择性使用抗菌素能发现线粒体蛋白质合成的位置。例如放线菌酮的存在可抑制细胞质核糖体，但某种线粒体蛋白照常合成，表明它是在线粒体核糖体上合成。有些线粒体蛋白在氯霉素存在下可以合成，而氯霉素是可抑制 mt 核糖体的，那么这些蛋白必定是在细胞质核糖体上合成的。用这种方法科学家们发现了酵母线粒体中细胞色素氧化酶有 7 个亚基，其中 4 个是在细胞质中合成的，其余 3 个是在线粒体中合成的(表 20－5)。

二、叶绿体基因组

叶绿体(Chloroplast)是一种只存在于绿色植物中的细胞器，主要负责光合作用。和线粒体相似，叶绿体也含有自己的基因组，但我们对叶绿体基因组的了解不如线粒体那样多。

表 20－5　酵母线粒体中合成的酶亚基

酶	分子量	亚基	
	(kDa)	线粒体中合成	细胞质中合成
ATPase(对寡霉素敏感)	340	ATPase 6、8、9(FO 膜因子)	ATP 1、2、3、4、7(F1 ATPase)
细胞色素 c 氧化酶	137	CO 1、2、3	CO 4、5、6、7
细胞色素 bc1 复合物	160	细胞色素 b 脱辅基蛋白	6 个亚基

参考 B. Lewin: Genes V，1994，Table 25.2

1. 结构

叶绿体基因组在很多方面与线粒体基因组的结构相似。表 20－1 表明了叶绿体 DNA(cpDNA)的某些特点。cp DNA 是双链环状,缺乏组蛋白和超螺旋。cp DNA 中的 GC 含量与核 DNA 及 mtDNA 有 很大的不同(表 20－1)。因此可用 CsCl 密度梯度离心来分离 cp DNA。

每个叶绿体中 cpDNA 的拷贝数随着物种的不同而不同。但都是多拷贝的。这些拷贝位于类核区。例如甜菜的叶细胞中每个类核体有 4～8 个拷贝的 cpDNA,而每个叶绿体有 4～18 个类核体,每个细胞中约有 40 叶绿体。每个细胞总共有约 6000 拷贝 cpDNA 分子。在衣藻中(*chlamydomonas*)(单细胞生物)一个叶绿体含有 500～1500 cpDNA 分子。

在很多植物中 cpDNA 长 40～45μm, 约是动物 mtDNA 的 8～9 倍, 一般在 121～155kb 之间。大的为 195kb(如衣藻 *chlamydomonas reinhardi*),小的也有 85kb。

2. 复制

cpDNA 的复制是半保留复制的过程,可能与线粒体相似。叶绿体特异性酶被用于此复制过程,这些酶不同于核和线粒体的复制酶系。大部分 cp 酶是由核基因编码,合成后转运到叶绿体中。叶绿体本身的生长和分裂以和线粒体相同的方式进行。

地钱(*Liverwort*)、烟草和水稻(*Oryza sativa*)叶绿体全序列分析(表 20－6)表明 cpDNA 基因组成有以下特点(图 20－4)。①基因组由两个反向重复序列(IR)和一个短单拷贝序列(short single copy seguence, SSC)及一个长单拷贝序列(long single copy seguence, LSC)组成;②IRA 和 IRB 长各 10～24kb,编码相同,方向相反。③cpDNA 启动子和原核生物的相似,有的基因产生单顺反子的 mRNA,有的为多顺反子 mRNA;④尽管 cpDNA 大小各不相同,但基因组成是相似的,而且所有基因的数目几乎是相同的,它们大部分产物是类囊体的成分或和氧化还原反应有关(表 20－7);⑤其 tRNA 基因(IRA、IRB 上各有 7 个,LSC 上有 23 个,共 37 个)中有内含子存在,最长者达 2526bp,这和原核 tRNA 不同。有的内含子位于 D 环上,这和真核生物核 tRNA 内含子常位于反密码子环上也不相同;⑥所有叶绿体基因转录的 mRNA 都由叶绿体核糖体翻译。

表 20－6 叶绿体基因组的组成

	地钱	烟草	水稻
反向重复序列	10,058	25,339	20,798
短单拷贝序列	19,813	18,482	12,335
长单拷贝序列	81,095	86,684	80,592
总基因组	121,024	155,844	134,525

参考 B. Lewin: Genes Ⅵ, 1997, Table 24.3

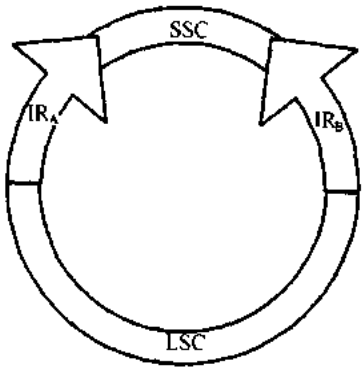


图 20－4 叶绿体基因组的结构。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 24－3)

并不是所有的叶绿体 DNA 都含有 IR，IR 上含有 4 种 rRNA 基因，根据它们排列的情况叶绿体可分为 3 类：I 类 IR 序列，4 种 rRNA 各有 2 个拷贝，对称分布在 IR 上（图 20－5）。cpDNA 也较大，如玉米、烟草、水稻、菠菜、地钱、衣藻（*C. yreinhardi*），大部分叶绿体都属此类。II 类：无反向重复 IR，而在 cpDNA 一侧 16S，23S 以正向串联重复的形式（各 3 个拷贝）排列。如少数低等植物，裸藻（*Euglena gracilis*）；III 类：无 IR 和 DR，rRNA 只有一拷贝，如豌豆（*Posum satirum*）等。这可能在进化的过程中 DNA 片段的重复和倒位而造成的。

3. 叶绿体的核糖体和蛋白质的合成

叶绿体蛋白合成所有的核糖体与线粒体和胞质核糖体完全不同。它们的沉降系数为 70S，由大小不等的两个亚基（50S 和 30S）组成，和原核生物的核糖体十分相似。

cp 核糖体的大亚基至少含有 2 个 rRNA 片段，每个片段一个拷贝：23S 和 5S。小亚基含有一个拷贝的 16S rRNA，很多核糖体蛋白是由核基因编码，在细胞质核糖体上合成的。大部分资料来自于对裸藻的（Eugleno）的研究，结果表明叶绿体产生 9 种核糖体蛋白，另外 12 种是在胞质中合成的。

叶绿中蛋白质合成的过程与原核生物相似，fMet-tRNA 用于各种蛋白质的起始，甲酰化反应是由位于叶绿体中的转

表 20－7　叶绿体基因组编码的各种蛋白

	基因	类型	拷贝
RNA－编码	16S	1	2
	23S	1	2
	4.5S	1	2
	5S	1	2
	tRNA	30	37
基因表达	r－蛋白	19	22
	RNA 聚合酶	3	3
	其他	2	2
类囊体膜	光系统 I	2	2
	光系统 II	7	7
	细胞色素 b/f	3	3
	H^+－ATPase	6	6
其他	NADH 脱氢酶	6	6
	铁氧化还原酶	3	3
	核酮糖 BP 羧化酶	1	1
	未鉴定	29	29
	总计	110	124

参考 B. Lewin: Genes VI，1997，Table 24.2

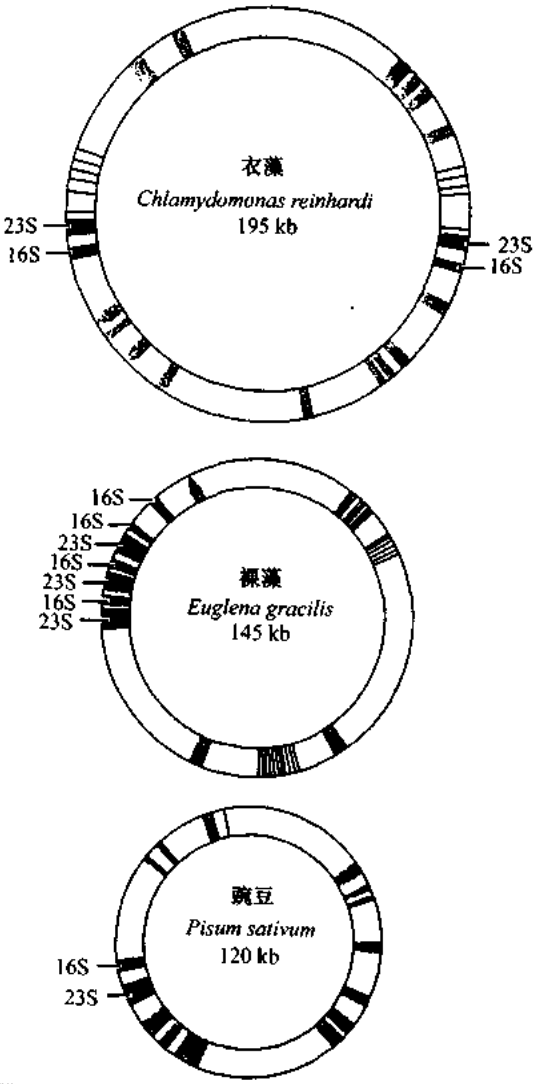


图 20－5　几种植物叶绿体基因组的组成。（a）衣藻（*Chlamydomonas reinhard*）的叶绿体；（b）裸藻（*Euglena gracilis*）的叶绿体；（c）豌豆（*Pisum sativum*）的叶绿体。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 21－13）

甲酰酶催化的。叶绿体中所用 IFs 和 EFs 和细胞质中蛋白质合成系统是不同的。

cp 核糖体像线粒体核糖体一样对氯霉素（可抑制细胞质核糖体）具有抗性，但对所有阻断原核蛋白合成的抑制剂都是敏感的。叶绿体中的核糖体本身也分为游离的和膜结合的两种。游离核糖体合成核酮糖二磷酸羧化酶（ribulose bisphosphate decarboxylase）的大亚基，而膜结合核糖体合成一种在光合作用中起重要作用的疏水蛋白。这种蛋白定位于叶绿体的膜中。

第二节　核外遗传

一、核外遗传的特点

位于细胞器的基因其遗传的规律不同于核基因，所以称为**非孟德尔遗传**（non - Mendelian inheritance）。这是由于细胞器的分离规律不同于染色体分离体规律所致。

核外遗传有四项主要特点：①正反交的结果不同。核外基因通常显示逐代出现**单亲遗传**（uniparental inheritance）的现象，即所有的后代不论雌雄都只有一个亲体的某一表型（此和性连锁是不同的）。最典型的是高等真核生物中，仅母亲的性状在后代中得到表达，这种现象称**母体遗传**（maternal inheritance）。这是由雌性配子—卵中含有大量的细胞质，而雄性配子—精子几乎不含细胞质，因此合子的细胞质基本上来自母本，而细胞质中含有大量的细胞器，如线粒体和叶绿体等。有的细胞器带有核外基因，形成了母体遗传。核基因的正反交第一代结果相同都显示显性亲本性状，性 - 连锁基因正反交第一代结果不同；性状和性染色的组合有关；核外基因正反交第一代结果不同；但 F_1 都显示母本的情况，和显隐性及性染色体的组合无关。②不出现分离比，即非孟德尔式遗传；③ 母本的表型决定了所有 F_1 代的表型；④遗传物质在细胞器上，不受核移植的影响；⑤不能进行遗传作图。

1. 叶绿体系统的核外遗传

第一个非孟德尔遗传的例子是在 1909 年由 Carl Correns 发现的。他是孟德尔原理重新发现的三个学者之一。他研究了大量显花植物的杂色叶片，这些表型中很多都显示了典型的孟德尔遗传。但他发现了一种例外的情况，在紫茉莉（*Mirabilis jalapa*）中有一种品系在其茎和叶上出现白、绿相间的绿白斑（albomaculata）（图 20 - 6）。以不同表型枝条上的花朵相互受粉产生种子后，其后代的表型完全取决于结种子的枝条（♀），而和采集花粉的枝条表型无关（表 20 - 8）。

结果显示了两个特点：①正反交的结果不同，例如：白色♀ × 绿色♂ 产生的结果不同于绿色♀ × 白色♂，也不出现孟德尔分离比。在核基因遗传中，只有性连锁基因正反交的结果不同，但用性连锁又不能解释此结果。②后代表型和母本表型一致，和父本表型无关。父本传给后代的遗传信息为零，这种方式就称为**母体遗传**。虽然白化的后代由于缺乏叶绿体而不能长期存活，但其他的后代能够幸存，可进行杂交。

如何来解释母体遗传呢？高等植物之所以能产生绿色是由于在叶绿体中存在着绿色的色素——叶绿素。绿色紫茉莉是含有正常的叶绿体成分，白色的枝条是含有无色的叶绿体，又称为**白色体**（leukoplasts），它缺乏叶绿素，故不能进行光合作用。从分子水

表 20－8　紫茉莉花斑植物杂交的结果

母本枝条表型	父本枝条表型	后代表型
白色	白色	白色
	绿色	白色
	花斑	白色
绿色	白色	绿色
	绿色	绿色
	花斑	绿色
花斑	白色	花斑，绿色，白色
	绿色	花斑，绿色，白色
	花斑	花斑，绿色，白色

Peter J. Russell: Genetics, 3rd ed., 1992, Table 21.5

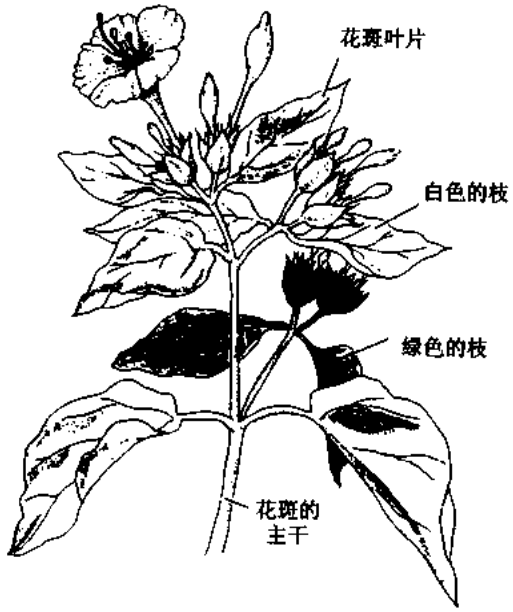


图 20－6　紫茉莉(*Mirabilis jalapa*)的花斑叶片。在同一植株上有的嫩枝是绿的，有的嫩枝是白的，有的嫩枝是花斑的，而在任何一个嫩枝上都有花。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 21－17)

平上来说这是由于其 cpDNA 发生了基因突变。在植物生长的过程中，叶绿体和白色体这两种细胞器可以分离，因此特殊的细胞及其后代细胞有的仅有叶绿体，而有的仅有白色体，或者二者兼有。绿、白花斑技物中白化的枝叶是来自仅有白色体的细胞，绿色的枝叶是来自有叶绿体的细胞，而具有花斑的枝叶是来自含有叶绿体和白色体的细胞。绿色枝条上的叶具有叶绿体，通过母体的遗传可将这些叶绿体传传递给后代，同样白色和花斑的表型也传给后代(图 20－7)。以上过程可描述为"**有丝分裂分离**"，但不是核基因的有丝分裂分离，而是用一个新的术语**暗盒**(black box)**学说**来描绘。在此暗盒中细胞器的基因组进行分离和重组。我们应称之为**细胞质分离和重组**(cytoplasmic segregation and recombination, CSAR)。

这一简单的模型有三项假设：①花粉的细胞质中不含有遗传信息(即没有叶绿体或白色体)，卵细胞比花粉要大得多。受精时也并不是整个的花粉和卵细胞融

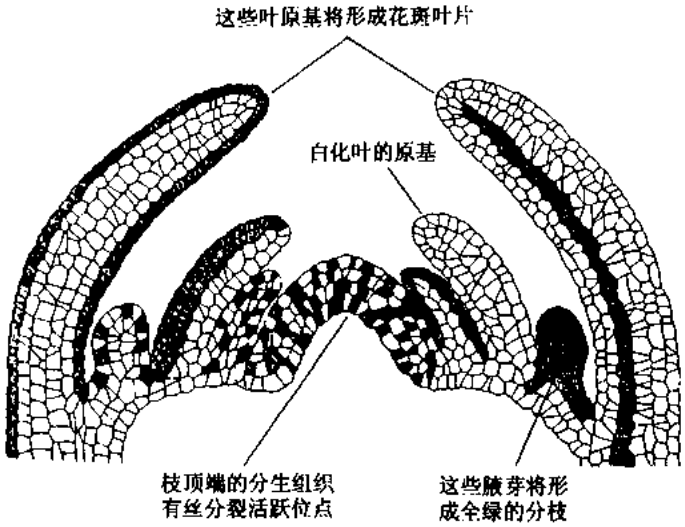


图 20－7　在含有绿色和白色两种叶绿体的植物中绿色或白色的细胞系建立于分生组织(生长点)。这种植物的卵细胞也可能含有两种类型的叶绿体。受精后经有丝分裂细胞质分离,形成纯的细胞系。这样植株任何部分的颜色都区决于含有何种类型的叶绿体。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 21－4)

合，而雄核和营养核经花粉管进入胚囊中，这样合子中的核外遗传基本取决于卵细胞，而和花粉无关；②叶绿体基因组通过质体的生长和分裂，进行自我复制。野生型和突变型 cpDNA 分子随机地进入新的质体中，因此纯系可能由混合系产生的；③质体分离，进入子细胞是随机的，有的子细胞接受了叶绿体，有的接受了白色体，有的二者都接受了。

2. 真菌的细胞质遗传

小菌落脉孢菌 (*Neurospora*) 的遗传　1952 年 Mary Mitchell 分离了脉胞菌的突变品系。称为**小菌落**(poky)。此突变体在很多方面都不同于野生型真菌。它生长慢,并表现出母体遗传的特点,细胞色素的总量异常。细胞色素是线粒体的电子传递体蛋白,是生物氧化酶类的辅助因子。生物通过氧化产生能量需要它。像大部分生物一样,野生型脉胞菌含有 3 种主要的细胞色素:a、b 和 c,但在 poky 小菌落品系中缺乏细胞色素 a 或 b,而含有 c。

在单倍体中如何论证母体遗传呢?此是可以通过野生型和小菌落品系的正反交来说明(图 20－8)。Mitchell 进行了下例的杂交实验:

poky 小菌落♀ × 野生型♂ →后代全部为 poky 小菌落。

野生型♀ × poky 小菌落♂ →后代全部为野生型。

在这两组杂交中,后代表型都和母体相同。核基因的遗传就与此不同了。两个亲本无论哪一个是 ad^+ 或 ad^-,其后代子囊中,有半数子囊胞为 ad^+,半数为 ad^-,显示了正常的孟德尔分离比 1∶1。所有的 poky 小菌落表型的后代都像原来的 poky 小菌落品系,在杂交中作为母本时就把 poky 小菌落传递给许多后代,但并不呈现一定的分离比。因而不可能是核基因突变之故。这种突变就称为**核外突变**(extranuclear mutation)或**细胞质突变**(cytoplasmic mutation)。

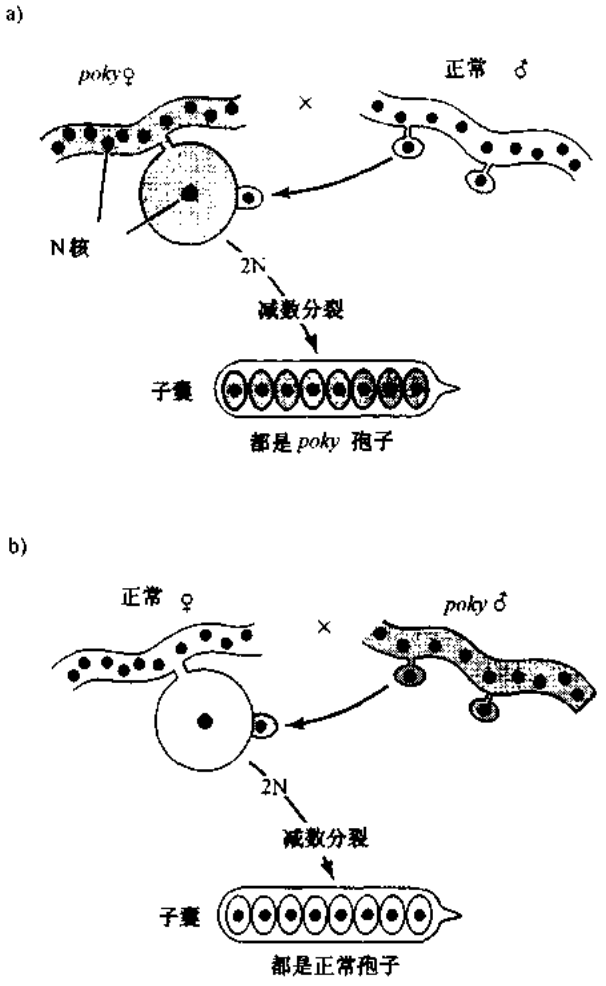


图 20－8　*poky* 小菌落和野生型链孢霉互交的结果(a)*poky* ♀ × 野生型 ♂；(b)野生型 ♀ × *poky* ♂。
(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 21－19)

现在进一步要问细胞质中哪一部分发生了突变呢？从突变表型看来和线粒体有关。例如细胞生长慢，表明缺乏 ATP 能量，在正常情况下，ATP 是由线粒体产生的，另外在突变体中细胞色素总量异常，而细胞色素也是位于线粒体膜上的。这使遗传学家得出结论，这种突变和线粒体有关，并开展了系列的研究，设计了很多有趣的实验来研究线粒体的自主性。

David Luck 用放射性胆碱（膜的一种成分）来标记线粒体，然后让它在非标记的介质中分裂并进行放射自显影，他发现，经几次分裂后，放射标记仍分布在线粒体中。Luck 得出结论认为，线粒体是通过原来存在的线粒体的分裂来增殖的。如果线粒体是重新合成的话，那么产生的线粒体群体就不会被标记，而原来的线粒体应保持较强的标记。

1965 年 Edward Tatum 领导的研究小组从与 poky 品系相同的异常品系脉胞菌中抽提纯化了线粒体，用一个极精细的针头和注射器将这种抽提的线粒体注入到野生型的受体细胞中，这些受体细胞经多次培养，其“异常”表型依然存在。线粒体的移植，传递了异常表型的遗传决定。推测这个涉及异常表型的基因是位于线粒体上。这一推测增加了在脉胞菌和其他物种的线粒体上存在基因的可信性。

3. 异核体测验

母体遗传是认识细胞器基本遗传的一个准则，另一些诊断实验应用于丝状真菌如脉胞菌和曲霉。从原则上说，这些测验能应用于和**异核体**(heterokaryors)有关的其他系统。异核体是由假定的核外突变（生长慢的突变）和带有已知的核基因突变之间产生的。如果核突变的表型在与缓慢生长突变表型的联合中形成异核体而得到恢复。那么可以假设生长突变来自细胞器突变。这是由于在异核体中一般不会产生二倍体，因此表型的恢复不大可能是核基因显隐性关系或重组所致。缓慢生长的表型一定是由细胞质含有的因子所决定的(图 20－9)。

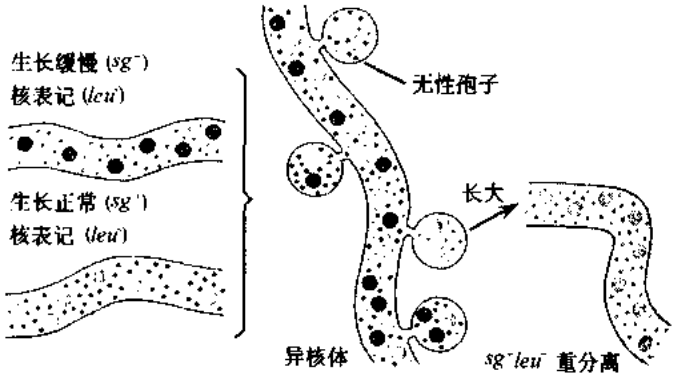


图 20－9 异核体测验被用于检测丝状真菌的核外遗传。一个带有核外突变的品系(sg^-，导致生长缓慢)与另一核突变品系(leu^-)联合，形成了异核体。此异核体培养可呈现 leu^- 表型。若某些后代出现 sg^- 表型，则 sg 很可能是核外基因。因在异核体中未发生核重组，所以 sg^- 表型必定和细胞质相联关。

(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 21－7)

二、酵母小菌落(petite)突变

酵母是单个细胞生长的，而不是像脉胞菌那样的以菌丝体生长的。因此在固体培养基上形成不连续的菌落。每个菌落是由数以千计的单个细胞在一起组成的。它可在有氧或无氧的条件下生长。在无氧条件下酵母通过发酵代谢来获得能量供给细胞生长和细胞的代谢，在发酵中并不涉及线粒体。

1. 小菌落的特点

1940 年后期 Boris Ephrussi 和他的同事研究了酵母细胞在固体培养基上的生长要么是通过需氧呼吸，要么通过发酵。偶然的机会，他们注意到一种比野生型菌落小得多的菌落，由于 Ephrussi 是法国人，所以称小菌落为 petites(法语“小”的意思)，而野生型菌落就称为 grandes (法语“大”的意思。)他发现小菌落并不是因为细胞小，而是因为其生长率明显比野生型低得多。

生化分析表明，小菌落主要是不能执行需氧呼吸，因此必须从发酵中获得能量，发酵是一个相对无效的过程。

在正常的细胞群体中约有 0.1%～1%的细胞会自发地变成小菌落。在碱基插入剂(intercalating agents)如溴化乙锭(EB)(见第 21 章)存在时 100%的细胞都变成了小

菌落。酵母小菌落系统常用于研究核外遗传，这是由于在酵母中这种突变很多，而且酵母缺乏线粒体的功能仍能存活和生长。这种引起呼吸不足的突变在酵母中可自发产生。在能维持发酵的培养基上，小菌落生长比大菌落慢一些，在维持需氧呼吸的培养基上小菌落是不能生长的。

2. 核基因突变型，中性型及抑制型小菌落突变

现在让我们回到小菌落遗传学上。有的小菌落是核基因发生了突变，这些核基因编码了某些线粒体蛋白的亚基。这种突变称为 *pet*⁻。当将 *pet*⁻ 与野生型 *pet*⁺ 杂交时产生的二倍体 *pet*⁺ / *pet*⁻ 是大菌落。这个二倍体细胞经减数分裂产生的四分体中 2 个是大菌落(*pet*⁺)，2 个是小菌落(*pet*⁻)，分离比为 2 : 2(表 20 - 9)。这个结果是符合孟德尔比，因此这种核突变型小菌落我们不再进一步讨论。这种核基因突变型的小菌落的发生频率也核外小菌落要低得多。

另一种类型的突变种为中性型小菌落(*neutral petite*)，以[rho⁻N]来表示。当中性型突变与野生型[rho⁺N]杂交产生了二倍体[rho⁺N]/[rho⁻N]，表型正常，减数分裂后产

表 20 - 9　酵母小菌落的类型和特点

小菌落类型	符号	产生原因	与野生型杂交	二倍体表型	四分体的基因型，表型和分离比
分离型	*pet*⁻	核基因突变	*pet*⁻ × *pet*⁺ →	*pet*⁻ / *pet*⁺(正常)	→小菌落(*pet*⁻)：大菌落 = 2:2
中性型	[*P*⁻ *N*]	mtDNA 丢失	[*P*⁻ *N*] × [*P*⁺ *N*] →	[*P*⁻ *N*]/[*P*⁺ *N*]→(正常)	→全为大菌落[*P*⁺ *N*]
抑制型	[*P*⁻ *S*]	mtDNA 突变	[*P*⁻ *S*] × [*P*⁺ *S*]→	[*P*⁻ *S*]/[*P*⁺ *S*] →(部分呼吸)	→小菌落[*P*⁻ *S*](迅速形成孢子) →小菌落[*P*⁻ *S*](长时间分裂)比例不等(1% ~ 99%)

生的四分体，其小菌落和大菌落之比为 0 : 4。显示出典型的单亲遗传，所有的后代只显示出一个亲本的表型。但我们不能把这一现象归结于母体遗传。因为两个单倍体细胞融合的二倍体双方细胞质的贡献是相同的。

中性型小菌落的特点是什么呢？前面说过当我们用插入剂像 EB 来处理酵母，很容易诱导小菌落的产生。而用较长的时间处理可以获得中性型小菌落。在中性型小菌落中线粒体遗传物质的研究提示了一个明显的特点：其 mtDNA　99% ~ 100% 丢失了。中性小菌落是没有线粒体功能的，它们能活下来是由于可以在细胞质中进行发酵过程。在与带有正常线粒体的野生型酵母的杂交中产生的后代都有线粒体，"小菌落"一代以后就消失了，这种小菌落是一种中性突变，并不影响到正常的表型(表 20 - 9)。

抑制型小菌落不同于前两种小菌落，大部分[rho⁻]突变是属于抑制型的[rho - S]。和中性型及核突变型相似，[rho⁻S]小菌落是不能合成线粒体蛋白，但和它们不同的是在它与野生型杂交形成的二倍体中它能发挥作用。

当[rho⁺]/[rho⁻S]二倍体形成时，它具有呼吸性能，比野生型弱，比小菌落强，居二者之间，若在二倍体产生之后立即形成胞子的话，结果四分体中，小菌落：正常菌落 = 4 : 0，即全部后代为小菌落。若二倍体的有丝分裂的时间加长（图 20 - 10）的话，后代群体中小菌落的比例不等，变化幅度很大，在 1 ~ 99% 之间，取决于特殊的抑制型小菌落品系。在这个群体中任何正常亲体的胞子都是正常的，即小菌落：大菌落 = 0 : 4(表 20 - 9)。

这种抑制型突变是由于 mtDNA 发生改变所致，开始它们缺失了部分的 mtDNA，而

后来通过某些机制使得序列不再缺失，而且发生重复，直到 mtDNA 的总量正常为止，从而序列得到了恢复。在这个过程中有时 mtDNA 会发生重排。由于线粒体中蛋白编码基因广泛地被播散，mtDNA 任何重要的缺失和重排大概都会导致线粒体蛋白合成的缺陷，特别是会影响到需氧呼吸的酶类（表 20－9）。

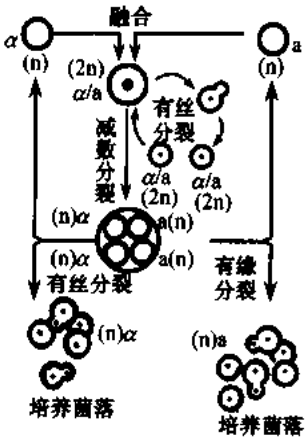
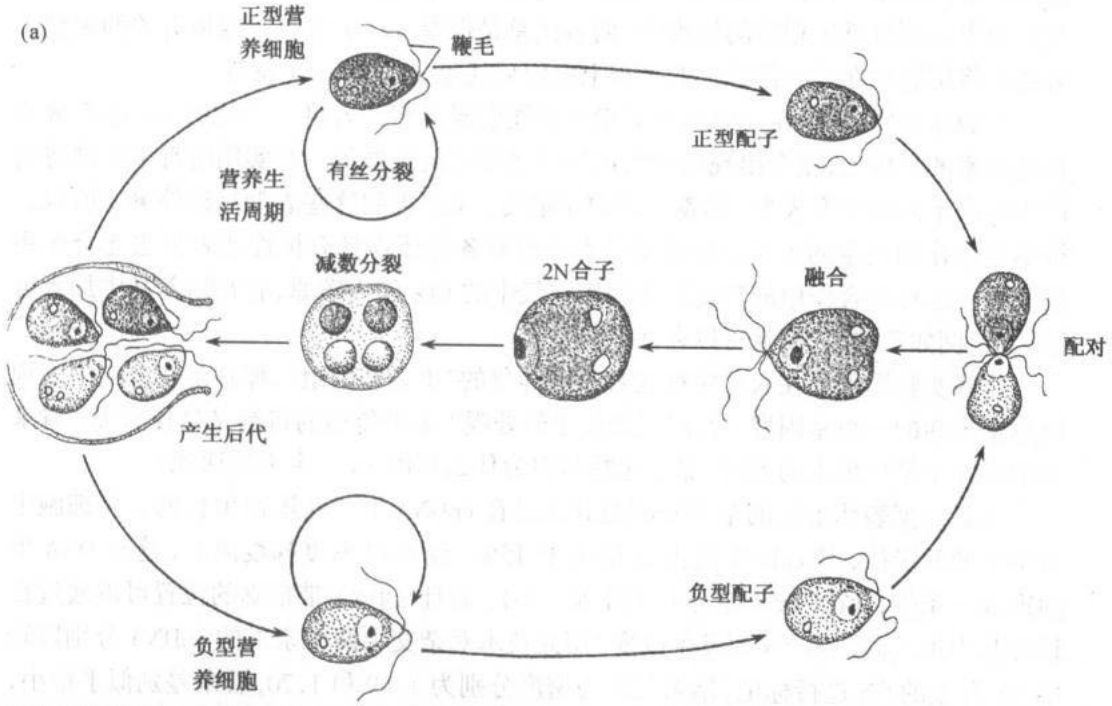


图 20－10　酵母的生活史。
（仿 A.J.F.Griffiths *et al*.：An Introduction to Genetics Analysis，4th ed.，1993，Fig.21－13）

三、衣藻的核外遗传

1954 年 Ruth Sager 和她的同事们研究了单细胞藻类——衣藻（*chlamydomonas reinhardi*）对抗菌素的抗性。衣藻是能游动的单倍体生物，它有两根鞭毛和一个含有多个 cpDNA 的叶绿体。图 20－11表



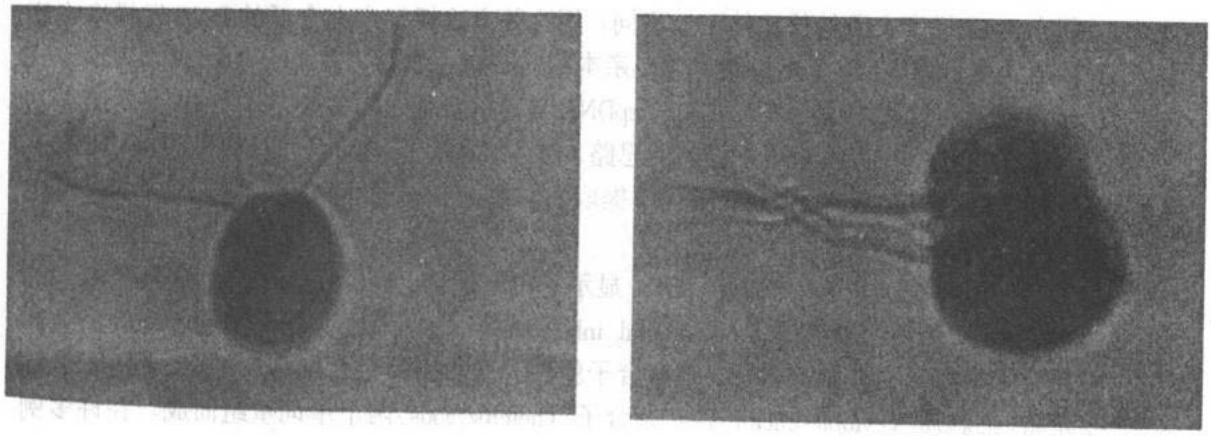


图 20－11　(a)单细胞绿藻生活史；(b)单个衣藻的光镜照片；(c)＋和－的衣藻融合。
（引自 Russell，P.J.：Genetics，3rd ed.，1992，Fig.6－2）

明它们的生活周期。它们有两种交配型 mt^+ 和 mt^-，通常简写成 + 和 −，是由一个座位上的等位基因控制的。属有性生殖，或**异宗配合**(heterothrothallic)的。合子是由两个大小相同的细胞融合而形成，它们含的细胞质的量也是相等的，但交配型不同。厚壁的胞囊发育成合子，经减数分裂后产生了四个单倍体的后代细胞，由于交配型是由核基因决定的，所以后代"+"、"−"交配型的比为 2 : 2。

对这种生物的特性进行了系统地研究，鉴别了大量的性状表现出核外遗传的一些模式，其中有些核外遗传的情况和抗性有关或依赖于抗性，通过改变核糖体的结构和功能来影响叶绿蛋白的合成。

有一个以核外遗传的性状是对红霉素（ery^r）(erythromycin resistance) 和链霉素的抗性(sm^r)。野生型的衣藻细胞是链霉素 − 敏感型(sm^s)。若我们进行下列杂交

$sm^r\ mt^+ \times sm^{-s}\ mt^- \rightarrow$ 全部后代(95%)sm^r

$sm^s\ mt^+ \times sm^r\ mt^- \rightarrow$ 全部后代(95%)sm^s

正反交结果不同，所有的后代(约 95%)表现了 mt^+ 亲本对抗菌素反应的表型，就像母体遗传现象一样，出现了单亲遗传(uniparental inheritance)(图 20 − 12)。但两个亲本对合子来说贡献了相等的细胞质。而后代总是保留 mt^+ 亲本与叶绿体有关的表型。可通过将细胞放在含有链霉素的平板上确定后代是 sm^r 或 sm^s 表型。

在做这个实验时 Sager 发现链霉素本身能引起突变，若将一个纯的 sm^s 品系放在含链霉素的平板上，就会出现相当大比例的链霉素抗性群落。其实用链霉素处理将会产生很多不同的突变表型，但都显示单亲遗传，从而推测这些表型是核外突变所致。链霉素的作用产生的大部分突变都是表现出对各种药物具有抗性或者失去光合作用的能力。这种光合作用的突变不能利用空气中的 CO_2 作为碳原，若在培养基中加入可溶性碳，例如乙酸，它们就可以存活。

这些实验揭示出在衣藻中存在着一种神秘的"单亲基因组"，那就是在杂交中表现出单亲传递的一组基因群。它们究竟位于何处呢？单亲传递的机制又是什么呢？看来交配型并不是生理上的差异，那么这些基因为什么只由 mt^+ 亲本传递呢？

实验证据表明上述的单亲基因组其实是在 cpDNA 上。在迅速生长的衣藻细胞中有单个的叶绿体，其 cpDNA 约占总 DNA 的 15%，经 CsCl 密度梯度离心，这种 DNA 单独形成一条带，和核 DNA 带分开（图 20 − 13），而且 cpDNA 带精确的位置可以通过在培养基中加入同位素 ^{15}N 而发生改变。用此技术对杂交中两个亲本的 cpDNA 分别以轻的 ^{14}N 和重的 ^{15}N 进行标记，结果其浮力密度分别为 1.69 和 1.70，看来差别似乎很小，但在 CsCl 梯度中所处位置的明显不同。用这种方法标记亲本合子的 DNA 经梯度密度离心再和亲本比较，就可以看出 mt^- 亲本的 cpDNA 丢失了(表 20 − 10)，或以某种方式破坏了。当然从 mt^- 亲本中丢失的 cpDNA 与其单亲基因的丢失是平行的。

其他的一些实验采用了相同的思路。例如有两个不同品系，它们的 cpDNA 带有不同的限制性酶切位点，从杂交后代中提取 cpDNA，经限制性酶处理，电泳后观察带型也能获得相同的结论。

前面我们也表明后代中仅 95% 显示了单亲遗传，那么另外的 5% 表现了双亲的核外特征，就称为**双亲遗传**(biparental inheritance)，表明双亲两种类型的叶绿体染色体在合子中都存在并具有活性。这些合子就称为**胞质基因杂合细胞**(cytohet)。这个名称是将"细胞质"(cytoplasmicall)和"杂合子"(heterozygous)两个单词重组而成。在许多例子中双亲合子的核外特征经有丝分裂分离成纯的类型，即 sm^r 和 sm^s 品系。这种现象是和合子中不同的 cpDNA 的分离有关。从而不同叶绿体进入子细胞，形成纯的类型。

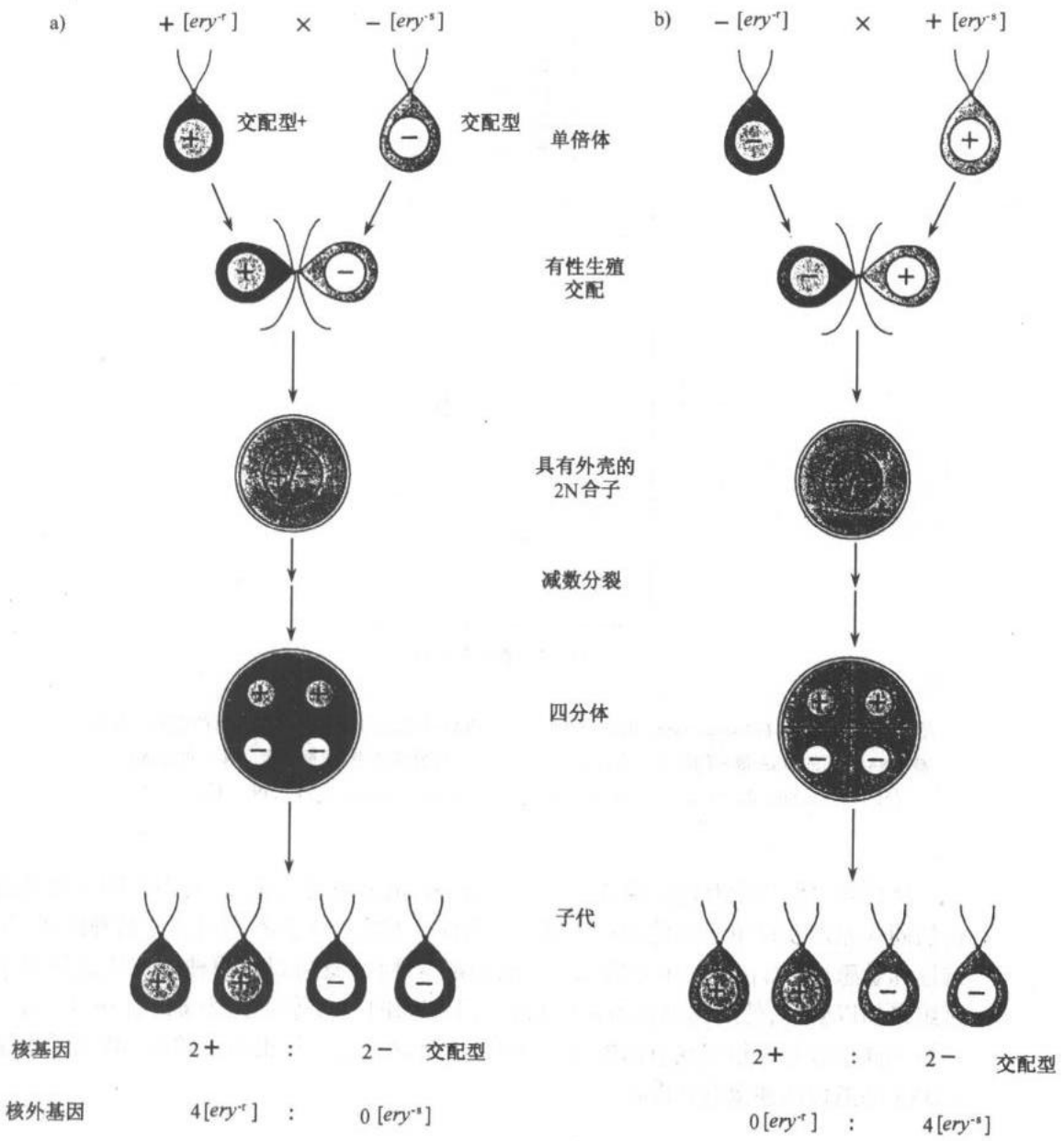


图 20－12 衣藻(*Chlamydomonas*)中的单亲遗传。(a) + [ery^{r}] × − [ery^{s}],95%的合子产生的四分体其核交配型基因以 2∶2 分离，+ 亲本所带的核外基因(ery^{r})以 4∶0 分离；(b)反交 − [ery^{r}] × + [ery^{s}]，95% 的合子产生的四分体以 2+ ∶2−，0[ery^{r}]∶4[ery^{r}]分离，+ 亲本的核外性状也表现为单亲遗传。

(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 21－22)

表 20－10 衣藻杂交后代的 cpDNA 的浮力密度

杂交	合子 cpDNA 的浮力密度
14N mt^{+} × 14N mt^{-}	1.69
15N mt^{+} × 15N mt^{-}	1.70
15N mt^{+} × 14N mt^{-}	1.70
14N mt^{+} × 15N mt^{-}	1.69

Ruth Sager: Cytoplasmic Gene and Organelles, Academic Press

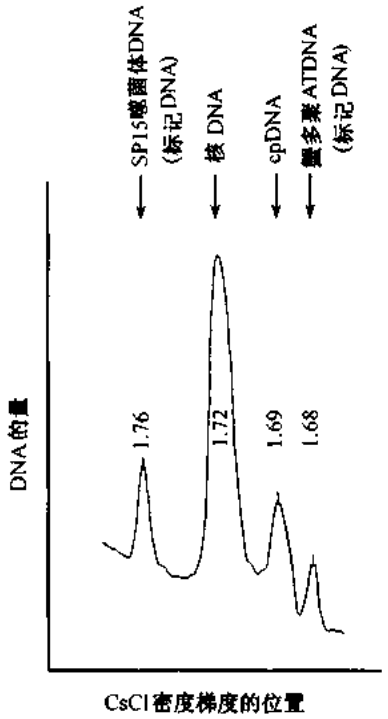


图 20－13　衣藻(*Chlamydomonas*)的叶绿体 DNA(cpDNA)可用 CsCl 密度梯度离心所产生的一条与核 DNA 分离的 DNA 带来的测定。在此实验中还会产生另外两条带,根据其密度梯度可加区别。
(转引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 21－11)

遗传学家们将带有由叶绿体控制的不同特征的品系进行杂交,与对抗菌素抗性遗传的同样原理获得单亲遗传和双亲遗传的后代。双亲合子并不同时表达两种性状,如抗性和敏感性。而在后代中分别表达不同的亲本性状。这可以用两种 cpDNA 之间发生重组来加以解释,使后代染色体带有原来属于双亲的一对等位基因(*sm*^{s} 和 *sm*^{r})。Sager 和她的同事收集了很多这方面的重组资料,在此基础上计算出重组频率,作出了衣藻 cpDNA 的遗传图和限制性图谱。

第三节　感染遗传——杀伤性酵母

除了叶绿体,线粒体等细胞器以外,真核生物细胞质中的细菌或病毒的共生也显示出核外遗传的特点。一个例子就是酵母的杀伤型表型。有些品系的酵母分泌一种杀伤性的毒素,可以杀死对此毒素敏感的品系(杀伤品系本身对此毒素是具有免疫力的)。此杀伤表型是由于在细胞质中存在两种类型的病毒:L 型和 M 型(图 20－14a),二者在宿主细胞中产生有害的作用。

L 型病毒在外壳蛋白中含有 4.5kb 的双链 RNA 基因组,称为 **L-dsRNA**,编码 L 型和 M 型病毒蛋白和复制病毒 RNA 的聚合酶序列。因此两种病毒的外壳蛋都是由 L-dsRNA 编码的,若细胞中有 L 病毒才可能存在 M 型病毒,另外它还含有 2 个相同 1.8kb 的双链 RNA 基因组,称为 M-dsRNA。**M-dsRNA** 编码杀伤毒素蛋白,从细胞中分泌出去。相同的蛋白负责产生对杀伤细胞的免疫性。

敏感型酵母细胞能被杀伤型细胞所分泌出的毒素所杀死,敏感型酵母也有两种类

型(图 20－14b)。一种类型只有 L 型病毒,另一种类型两种病毒都不存在。这两种类型的细胞因不能产生毒素所以没有免疫力。

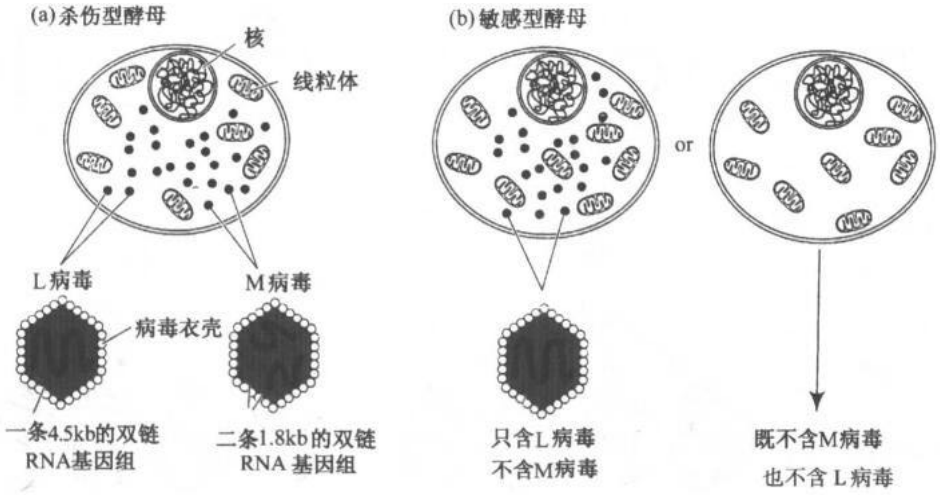


图 20－14 酵母中的杀伤现象。(a)杀伤型酵母含有 L 和 M 两种类型的病毒,这两种病毒都具有双链 RNA 基因组。L 链编码两种病毒的颗粒和复制所需的复制酶。M－链编码杀伤性毒素;(b)敏感型酵母可被杀伤性毒素杀死,这种酵母仅含有 L 病毒或既不含 M 病毒,也不含 L 病毒。

(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 21－24)

和大部分病毒不同,酵母的 L 和 M 病毒在细胞外不存在,因此敏感的品系也不会被细胞外的这两种病毒所感染。而病毒的传递是通过酵母在接合过程中细胞质的混合而进行的,所有的后代都继承了亲代的细胞中的病毒,产生了细胞质遗传的感染机制。例如杀伤型酵母和敏感型酵母接合产生的二倍体,由于有 L 和 M 病毒的存在此二倍体也成了杀伤型的。两种病毒分配到后代细胞中,使四分体都成为杀伤型的,这种情况也显示了核外遗传和单亲遗传的特点。

第四节 母体影响

母体遗传是否总是表现出核外遗传呢?回答是否定的。有些母体遗传的现象并不是由核外基因所控制。核基因的互交也会出现母体遗传的模式。早在 1923 年摩尔根的学生 Sturtevant 发现了一个很好的例子就是椎实螺(*Limnaea*)。这种小螺蛳的外壳有左旋和右旋两种。螺壳的这种旋转方向是单对核基因决定的。D 基因决定右旋,d 基因决定左旋。实验表明 D 对 d 为完全显性。而且螺壳的旋转表型总是由母本的基因型来决定 。图 20－15 表示右旋和左旋正反交的结果。它们的 F_1 代虽然核基因型相同,但表型却不同,都和母本的基因型一致,表现了母体遗传,但不符合核外遗传。当 F_1 代表型为右旋时,基因型为 D/d,互交时产生的后代核基因的分离比应为 $1(D/D):2(D/d):1\ (d/d)$,但表型都为右旋,和本身的基因型并不一致,而和母本的基因型是一致的。当 F_1 代表型为左旋时,基因型为 D/d,互交时产生的后代基因型也应按孟德尔法分离,但同样表型是一致的,和母体的表型(左旋)不同,而和其基因型一致。符合核外遗传。这种由母体的基因型决定后代表型的现象称为**母体影响**(maternal effect)。我们现在已经知道这种现象是由于母体的基因型决定了子代发育的起始剪接模式。图 20－15 表示这种分离比在真正的细胞器遗传中是不存在的。

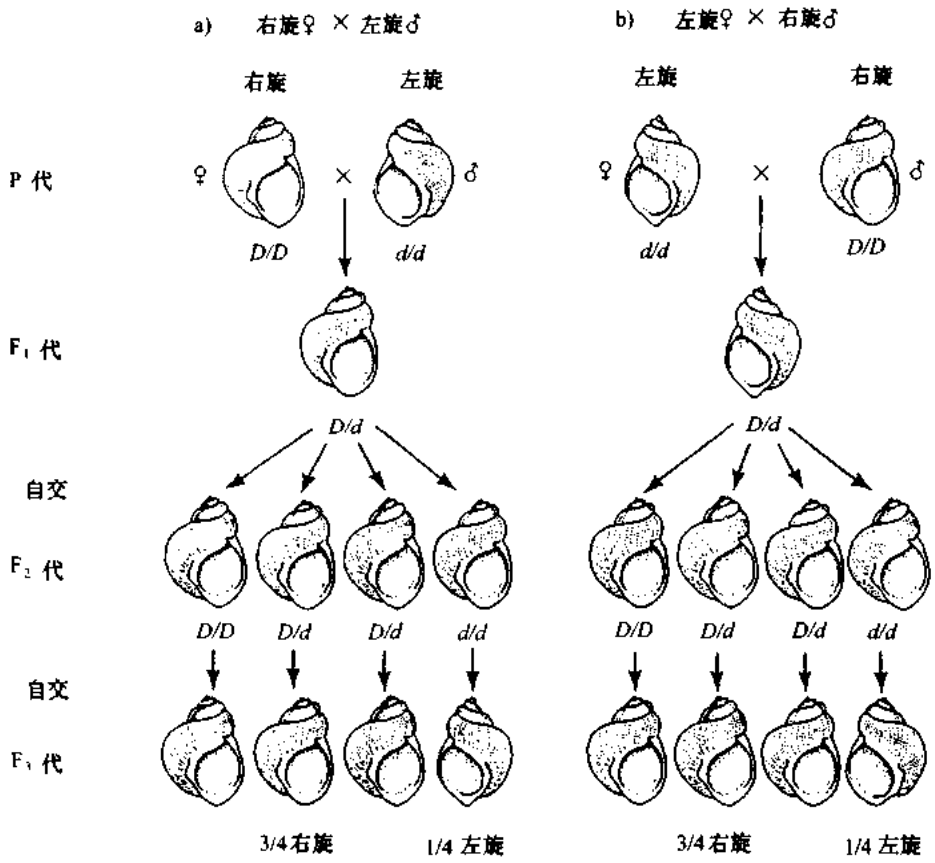


图 20－15　椎实螺螺壳旋转方向的遗传是母体影响的一个例子。(a)右旋(D/D)雌性和左旋(d/d)雄性之间的杂交;(b)左旋(d/d 雌性和右旋(D/D)雄性之间的杂交。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 21－15)

因此也不属于母体遗传。可能是母体核基因的产物影响了胚胎的发育，而不是细胞器中的遗传物质给了子代。

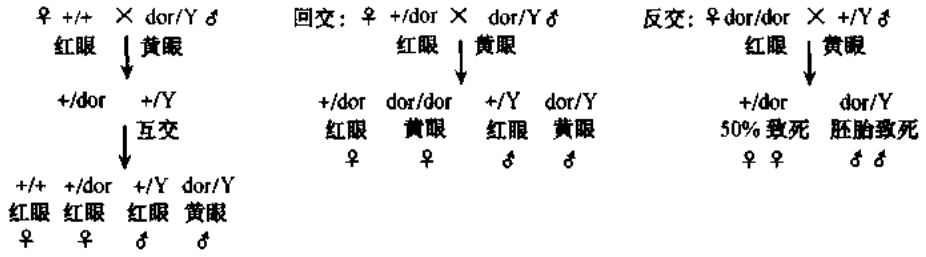


图 20－16　果蝇的伴性致死遗传模式。

另一个例子是果蝇的伴性致死或发育异常(图 20－16)。果蝇核基因 *dor* 突变会出现观桔黄色的眼色，此基因位于 x 染色体上，因此正反交的结果不同。正交和回交和伴性遗传的规律完全一致，但反交就大不相同，产生了胚胎死亡。然而它和致死基因又不相同，在正交和测交中 *dor*/ *y* 和 *dor*/ *dor* 黄蝇果蝇都能生存，但在反交中 *dor*/ + 都有 50%死亡，*dor*/ *y* 全部胚胎致死，这是什么原因呢?比较一下就会发现是和母体有关，可能是卵的细胞质中缺乏了什么。Garen 和 Gebring 抽取野生型细胞质注入了 *dor*/ *y* 和 *dor*/ *y* 的胚细胞质中，结果有 1/3 的胚细胞质发育到运动期，说明可能是卵母细胞中母

体的 mRNA 或其他某种产物所起的作用，而不可能是基因直接起作用。

第五节 真核细胞中的质粒

在细菌中存在着质粒是为大家熟知的现象（见第 10 章），在真核生物的细胞中也有质粒存在，只不过不是像细菌那样普遍。由于质粒一般和核染色体没有关系，因此表现出一种非孟德尔式遗传模式，和细胞器 DNA 的传递途径较为相似。

大部分真核质粒没有明显的表型效应，只能用一些分子水平的技术才能检测到它。搞得最清楚的就是酵母的 **2μ 环**（two－micron circle）。这个环状质粒可能位于核质。若一个含有此质粒的单倍体品系与另一个丢失了些质粒的品系接合，无论是有性还是无性后代都获得了此质粒。2μ 质粒在基因工程中可作为载体而显示了它的重要性。当将核 DNA 拼接成 2μ 载体并用于转化受体细胞时，插入片段被带到核中，在核中它可能通过各种重组机制参入到染色体中。

部分真核质粒是线粒体质粒。一个有趣的线粒体质粒是和玉米的细胞质雄性不育有关。在植物中雄性不育品系产生的花粉是没有功能的。当将雄性不育植物作为母本与正常可育的植物（作为父本）进行杂交时所产生的后代全部都是雄性不育的，这种遗传方式明显属于母体遗传。玉米中细胞质雄性不育是和两种线状质粒有关，它们是 **S1** 和 **S2**。它们位于线粒体中，附加在 mtDNA 上。虽然对这些质粒的分子特点已搞得较清楚，但还未能用精确地方法充分证明它们和雄性不育的关系。一个有趣的特点是 S1 和 S2 能与 mtDNA 重组。

玉米和其他农作物的雄性不育被用于杂交育种。在杂交育种中一个很大的问题就是如何防止自花受粉。以免干扰杂交种子产生。用人工去雄的办法是不可能在万顷大田中制种。而雄性不育就解决了这一问题。

人们常用的制种方法叫做**三系二区制种法**。所谓三系就是：

（1）雄性不育系，以符号（s）rfrf 来表示。（s）是指不育（sterility），用括号括起表示属于细胞质中的遗传因子，rf（restere fertile）是表示核内可使雄性不育得到恢复的一对等位基因。

（2）保持系：以符号（n）rfrf 来表示。这种品系的细胞质中没有雄性不育因子，是正常的，所以用（n）（normal）来表示。而核中 rf 是隐性的，不能使植物恢复可育。当它作为父本和不育系杂交后，雄性不育植株上结的穗子仍是雄性不育的，故称其为保持系。为了使植物充分地受粉和收集的方便，将雄性不育系和保持系种在一个区域，一行一行相间排列，使其充分得到杂交的机会，雄性不育株上的花粉是没有功能的，所以其腋间结的穗子一定是杂交后代，仍是雄性不育系，而保持系也只能是依靠自花受粉产生后代，所以此系统的穗子仍是保持系，这样就形成了“三系二区”中的一个区。用这种方法不仅使不育系和保持系的特点能稳定的遗传下去，解决了雄性不育传代的问题，同时还可以通过多次杂交，使保持系核基因所控制的优良性状传递给雄性不育后代，这样具有某种优良性状的不育系才能作为杂交育种中的母本。

（3）恢复系：（*N*）*RfRf*，这种品系不仅细胞质中无雄性不育因子，而且核中还存在恢复可育性的基因 *Rf*，*Rf* 对 *rf* 是显性，当恢复系和不育系杂交时产生的后代虽然细胞质中有不育因子存在，但核中的 *Rf* 基因可抑制其作用，这样就能产生杂合的雄性可育的种子：

♀（S）rfrf × ♂（N）RfRf→（S）Rfrf（雄性可育）

在育种中将恢复系和已具有某种优良性状的不育系种在同一地区，成行相间排列，这就是“二区”中的另一区，这样在不育系这一行玉米上所结的穗子就是我们所需要的杂交种子(图 20 - 17)。

另一个真核质粒的例子是在脉胞菌中的衰老品，在一般的培养基上不能生长，不

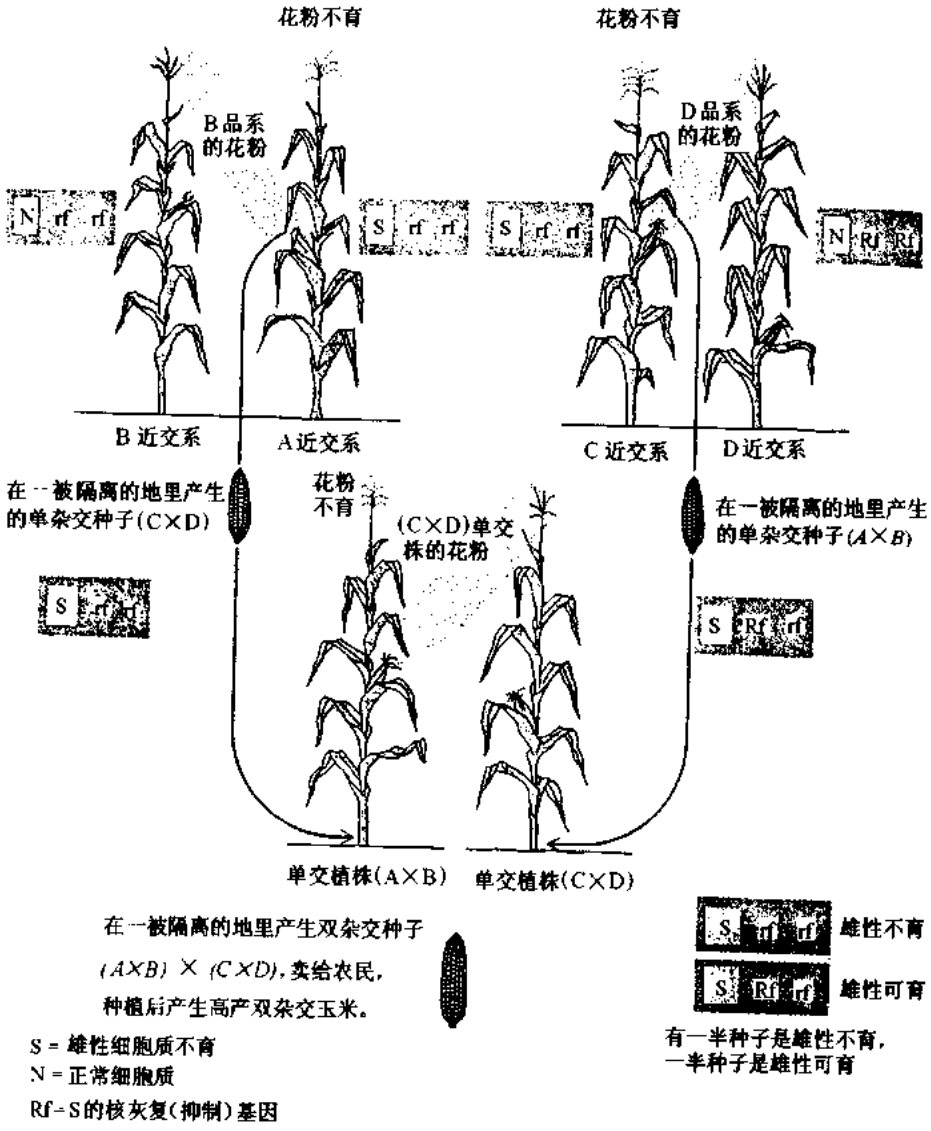


图 20 - 17　玉米的三系两区杂交育种法。(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 21 - 27)

久死亡。在夏威夷发现了这种群体，人们称这个品系为“Kalilo”，夏威夷语的意思就是“死亡之门”。Kalilo 品系仅在穿入培养基一定的深度时才能生长，然后就会死去，在未死前将这个品系进行杂交，其致死能力将以母体遗传的方式传递给后代。现已发现这种衰老的特征是由一长 9kb 的线粒体质粒决定的，这种质粒叫做 Kal DNA，它可以独立存在于细胞质中，而且是无害的。但一插入到 mtDNA 中就可引起逐步死亡。在死掉的细胞中，大部分 mtDNA 分子都有 Kal DNA 的插入(图 20 - 18)。推测这个品系的衰老和死亡可能由于插入 Kal DNA 的干扰或破坏坏了线粒体的正常功能。因此看来 Kal DNA 就像是个“分子寄生虫”。

值得注意的是在上述例子中这些质粒和玉米，脉胞菌的缺陷型表型有关联，但酵

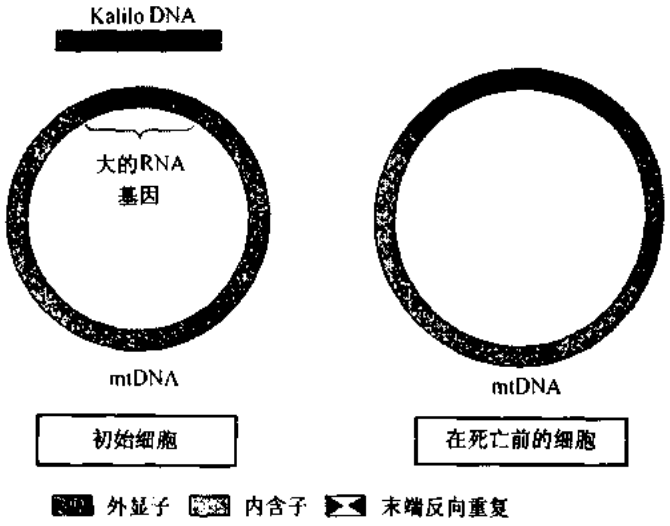


图 20－18　脉孢菌衰老品系是在其 mtDNA 上插入了一个线型质粒，称为 Kalilo DNA，它插入到内含子中，而其他的插入位点尚不清楚。（转引自 Griffiths A. J. F. et *al.*: An Introduction to Genetic Analysis, 5th ed., 1993. Fig. 21－28）

母的 2μ 质粒没有明显的表型效应。真核质粒的存在和进化地位仍然是个迷。但研究这些因子的结构和行为对于搞清遗传物质的基本特点可提供重要的思路。

第二十一章

基因突变

基本原理

突变是遗传物质中不是由于遗传重组产生的任何可遗传的改变。基因突变分为自发突变和诱发突变两类。

突变可能发生在染色体水平或基因水平，前者称为染色体畸变，后者称为基因突变。基因突变可以由于碱基对的置换或者一个或多个碱基的增加和减少而引起。

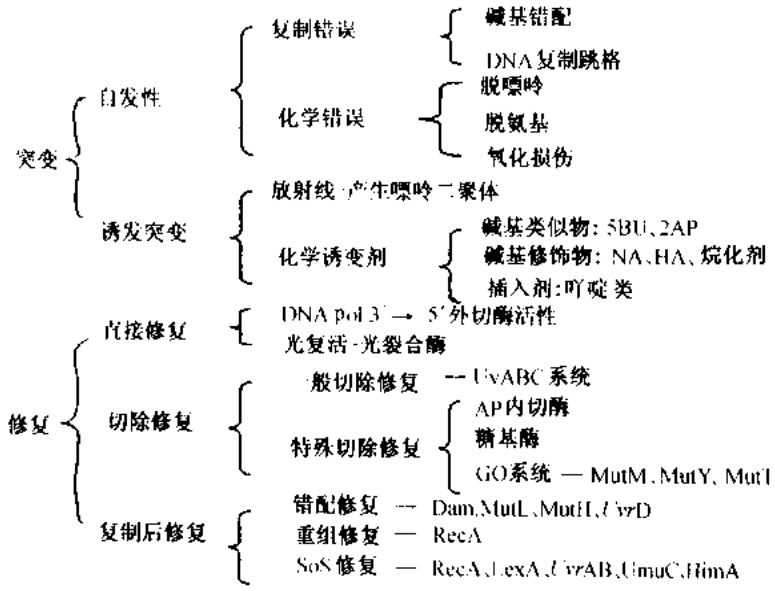
基因突变的结果取决于多种因素，例如错义突变可以导致氨基酸的置换，无义突变可以导致多肽合成的提前终止。

基因回复突变可以恢复原来的序列，也可以是在另一位点发生新的突变，结果使表型得到恢复。后者称为抑制，它们可以发生原来突变的密码子上，也可以发生在不同的密码子上。回复突变和原来的突变不在同一基因内时称为基因间抑制。这种抑制常涉及到 tRNA 反密码子的改变。

基因突变可由放射线所引起，放射线通过使染色体断裂产生遗传损伤。也可由化学物质和 DNA 的作用而产生，或者由于 DNA 之间异常链的形成而产生。一定种类的 DNA 损伤可能由放射线而引起，但不是染色体的断裂，它可被细胞中的修复系统所校正。

基因突变可以由某些化学物质所引起，这些化学物质称为化学诱变剂。现已知很多的化学诱变剂，它们诱变的机制是不同的。在 DNA 复制过程由于碱基类似物的取代，碱基的化学修饰以及碱基的插入和缺失都会引起突变。

在原核和真核细胞中都存在很多的修复系统，这些修复系统可恢复不同的 DNA 损伤。各种系统都是通过酶的作用来校正。有的系统是直接校正 DNA 损伤，有的是通过切除来进行修复(见下表)。



在绪论中我们谈到了遗传学是研究遗传和变异的一门学科。变异使得生物具有了多样性，造就了大千世界，变异使生物能通过自然选择而产生进化。变异是怎样发生的呢？遗传是以 DNA 的复制传递和表达为物质基础的，而变异同样是以 DNA 的各种改变为基础的，其中包括形成配子时不连锁基因的自由组合产生的重组型，基因突变，染

色体畸变以及 DNA 的重组和基因的重排，概括起来就是一句话，遗传信息发生改变。本章就着重阐明基因突变。

第一节　基因突变的类型

突变(mutation)是遗传物质中任何可检测的能遗传的改变，但不包括遗传重组。突变可以传递给子细胞，甚至延续给后代，从而产生了突变细胞或个体。对于一个多细胞生物来说，如果突变仅发生在体细胞中，那么这种突变是不会传递给后代的。这种类型的突变称为**体细胞突变**(somatic mutation)。但若突变发生在生殖细胞中，那么这种突变就能通过配子传递给下一代，在后代个体的体细胞和生殖细胞中产生同样的突变，这种突变就叫做**种系突变**(germ-line mutation)。由于遗传物质一般是 DNA，突变会影响到 DNA 的化学或物理的构成，它的复制，表型功能一个或多个碱基对序列会发生改变，例如增加、减少或置换碱基，颠倒序列或转移到新的位置上。

突变可以发生在染色体水平或者基因水平。染色体结构和数目的改变叫做**染色体畸变**(chromosomal aberration)，也称为**染色体突变**(chromosome mutation)(见第 22 章)。当染色体畸变涉及到基因组中染色体套数的改变称为**基因组突变** (genome mutation)，即整倍数改变。发生在基因水平的突变称为**基因突变**(gene mutation)，它涉及到基因的一个或多个序列的改变。包括一对或多对碱基对的替换，增加或缺失。由于 DNA 碱基对的改变引起的基因突变称为**点突变**(point mutation)

突变可以是自发的，但也可被某些诱变剂 (mutagen) 诱发。自然发生的突变是**自发突变** (spontaneous mutation) 。一些物理因素或化学试剂都能增加自发突变的频率。诱变剂处理所诱发的突变称为**诱发突变**(induced mutation)。这两种突变之间并没有本质的区别。突变表型显示出 DNA 改变所产生的后果，这是由于蛋白功能改变的结果。

有了以上这些背景，我们现在就可以介绍有关基因突变的一些术语。

表 21－1　突变的各种类型

- 突变的类型
  - 点突变 碱基替换
    - 转换：Py 与 Py，Pu 与 Pu 之间变换，多见
    - 颠换：Py 与 Pu 之间变换，少见
  - 移码突变
    - 插入 1、2 个碱基
    - 丢失 1、2 个碱基
    - → 发生移码
  - 缺失突变 —— 缺失大片段 DNA(十几到几千个碱基)

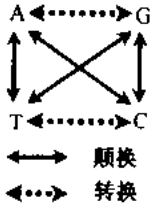


图 21－1　转换与颠换。

碱基对取代突变 (base-pair substitution mutation) 是指在基因中一个碱基对被另一碱基对所取代(表 21－1,2)。例如表 21－2 中 Tyr 的 DNA 密码子中 C-G 对变成 A：T 对。

转换(transition)是碱基替换中的一种类型，是指嘌呤与嘌呤之间，或嘧啶与嘧啶之间的替换(图 21－1)。

颠换 (transversion mutation) 是碱基替换中的另一种类型，是指嘌呤与嘧啶之间的替换(图 21－1)。

错义突变 (missense mutation) 是指 DNA 改变后 mRNA 中相

应密码子发生改变，编码另一种氨基酸（表 20－2），使蛋白质中的氨基酸发生取代，可能削弱此蛋白的功能，以至影响到突变体的表型。

突变体的表型是否易于检测取决于特殊氨基酸的取代。例如在人类中，β－珠蛋白基因中单个碱基的改变会导致 β-链中氨基酸的取代。如果个体是突变的纯合体，那就会出现镰形细胞贫血。但如果多肽链中氨基酸的取代不影响蛋白质的功能，那么是不会产生表型改变的。**无义突变**（nonsense mutation）是由 DNA 的碱基突变使 mRNA 中产生了无义密码子（表 20－2），翻译时在相应的位点会导致提前终止——平截或截

表 21－2 点突变的类型（以 Tyr 的密码子为例）

无义突变	同义突变	错义突变
DNA TAC→TAA，TAG	TAC → TAT	TAC→ TCC
↓ ↓ ↓ ↓	↓ ↓	↓ ↓
RNA UAC UAA UAG	UAC UAU	UAC UCC
↓ ↓ ↓ ↓	↓ ↓	↓ ↓
AA Tyr Och Amb	Tyr Tyr	Tyr Ser

短，产生一条不完整的多肽链，通常是没有功能的。

中性突变（neutral mutation）是在基因中有一对碱基对发生替换，引起的 mRNA 中密码子的改变，但多肽链中相应位点发生的氨基酸的取代并不影响蛋白质的功能，我们就称之为中性突变。例如密码子 ACG→AAG，那么导致了 Lys 取代了 Arg，这两种氨基酸都是碱性氨基酸，性质十分相似，所以蛋白的功能并不发生重大的改变。

沉默突变（silent mutation）是中性突变中的一种特殊情况。即在基因中碱基对发生替换，改变了 mRNA 的密码子，结果蛋白质中相应位点是发生了相同氨基酸的取代，由于氨基酸的序列未发生变化，所以蛋白质仍具有野生型的功能，实际上就是**同义突变**（表 20－2）。

移码突变（frameshift mutation）是由于基因中增加或减少碱基（改变的碱基数不得是 3 或 3 的倍数）所致。增加或减少 1～2 个碱基会使该位点后面的 RNA 序列发生移码，产生无功能的蛋白质。

回复突变（reverse mutation），根据突变表型和野生型相比较将点突变分成两类：一类是**正向突变**（forward mutation），其突变方向是从野生型向突变型；另一种是**回复突变**，其突变方向是从突变型向野生型

回复突变可使突变基因产生的无功能或有部分功能的多肽恢复部分或完全功能。例如当 DNA 碱基对发生改变，使其 mRNA 中相应的无义密码子得到回复，可以编码某种特殊的氨基酸，这个改变的密码子可以是原来野生型的相应密码子或者是其他氨基酸的密码子。

突变的作用还可以通过其他位点的突变而得到减少或校正，这便是前面已讨论过的**抑制突变**（suppressor mutation）（见第 15 章）。

第二节 突变的原因

我们已经阐叙了基因突变的类型，下面介绍自发突变和诱发突变的发生过程。

一、自发突变

自发突变(spontaneous mutation)是自然发生的,不存在人类的干扰。前面描述的各种类型的点突变都是自发的。长期以来遗转学家们认为自发突变是由环境中固有的诱变剂所产生的,如放射线和化学物质。但有证据表明虽然自发突变率非常低,但对于仅有的环境中固有诱变剂而言仍然是太高了。

人们试图对突变的发生进行定量。常使用两种不同的术语:**突变率和突变频率**。突变率(mutation rate)是指在单位时间内某种突变发生的概率,即每代每对核苷的突变概率数或每代每个基因的突变概率。突变频率的定义就不同了,它是指在一个细胞群体或个体中,某种突变发生的数目,即每10万个生物中发生突变体的数目,或每百万个体中突变的数目。

例如在果蝇中自发突变率大约为 10^{-4} ~ 10^{-5} 每代每个基因,在人类中为 10^{-4} ~ 4×10^{-6} 每代每个基因,而细菌是 10^{-5} ~ 10^{-7} 每代每个基因。多种生物特殊座位的自发突变的频率如表 21 - 3 所示。自发突变率受到生物遗传特征的影响,比如在雄性和雌

表 21 - 3　多种生物特殊基因座的自发突变频率

生物	性状	突变频率	生物	性状	突变频率
T2 噬菌体	快速溶菌 $r^+\rightarrow r^-$	7	小鼠	$a^+\rightarrow$非鼠色	0.97
	宿主范围 $h^+\rightarrow h^-$	0.001		$b^+\rightarrow$褐色	0.39
E. coli	ste^{-r}	0.00004		$c^+\rightarrow$白化	1.02
	对 T1 产生抗性	0.003		$d^+\rightarrow$淡化	1.25
	$leu^-\rightarrow leu^+$	0.00007		$In^+\rightarrow$灰色	0.80
	$arg^-\rightarrow arg^+$	0.0004		回复突变	0.27
	$trp^-\rightarrow trp^+$	0.006	中国仓鼠	氮鸟嘌呤抗性	0.0015
	$ara^+\rightarrow ara^-$	0.2		$glu\rightarrow glu^+$	0.014
沙门氏细菌	$his^+\rightarrow his$	0.2	人类	软骨发育不全	0.6 ~ 1.3
	$trp\rightarrow trp^+$	0.005		先天性无虹膜	0.3 ~ 0.5
肺炎双球菌	$pen^s\rightarrow pen^r$	0.01		萎缩性肌强直	0.8 ~ 1.1
粗糙脉孢菌	$ade^-\rightarrow ade^+$	0.0008 ~ 0.029		结节性(脑)硬化	0.4 ~ 1
	$ino^-\rightarrow ino^+$	0.001 ~ 0.010		亨廷顿舞蹈症	0.5
	肌醇等位基因(JH5202)	1.5		神经纤维病	5 ~ 10
黑腹果蝇	$y^+\rightarrow$黄色	12		肠多发性息肉	1.3
	$bw^+\rightarrow$褐色	3		成骨不全	0.7 ~ 1.3
	$e^+\rightarrow$檀色	2		佩-格尔异常	1.7 ~ 2.3
	$ey^+\rightarrow$无眼	6		视网膜母细胞瘤	0.5 ~ 1.2
玉米	$Wx\rightarrow wx^-$	0.00			
	$Sh\rightarrow sh$	0.12			
	$C\rightarrow c$	0.23			
	甜 $Su\rightarrow su$	0.24			
	紫 $Pr\rightarrow pr$	1.10			
	$I\rightarrow i$	110.26			
	$R^r\rightarrow r^r$	49.20			

转摘自 Peter J. Russell: Genetics, 3rd ed., 1992, Table 23.5

性果蝇中相同的性状其突变率不同。

自发突变可能由很多因素中的一种所引起，包括 DNA 复制中的错误，及 DNA 自发的化学改变。自发突变也可能由于转座因子的移动而引起(见 23 章)。

1. DNA 复制错误

在 DNA 复制时可能产生碱基的错配，如 A-C 配对。当带有 A-C 错配的 DNA 重新复制时，产生的两条子链中，一条子链双螺旋在错配的位置上形成 G-C 时，而另一条子链的双旋在相应位点将形成 A-T 对。这样就产生了碱基对的转换。由于碱基它们本身存在着交替的化学结构，称为**互变异构体**(tautomers)，所以也能形成错误的碱配对。当碱

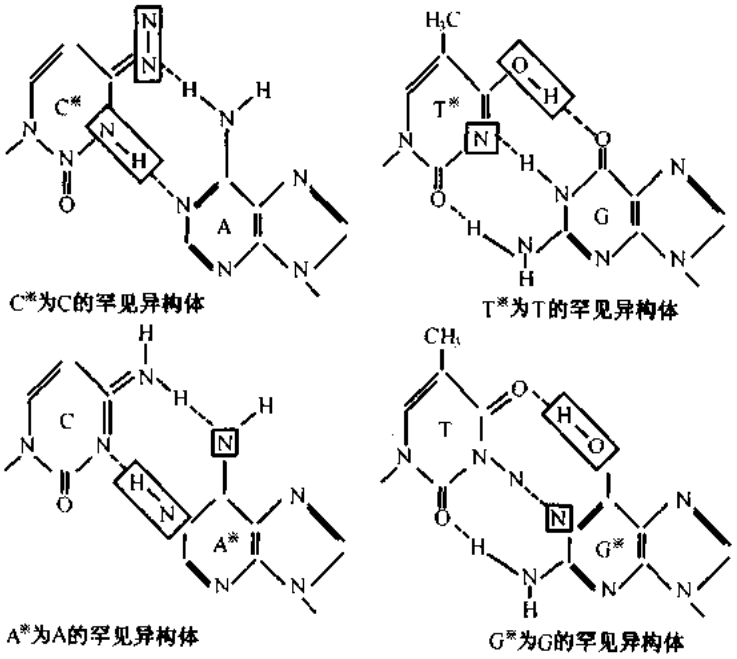


图 21－2　碱基的罕见异构体形成的错误碱基配对。(仿 Peter J. Russell: Genetics, 3 rd ed., 1992, Fig. 18－5)

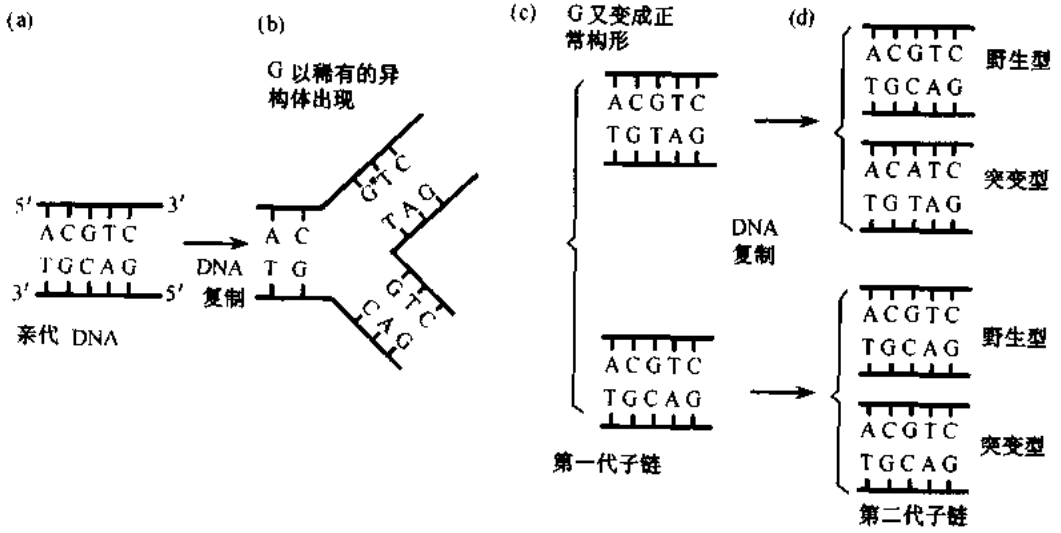


图 21－3　在 DNA 碱基中互变异构移位产生的突变。(仿 Peter J. Russell: Genetics, 3rd ed., 1992, Fig. 18－6)

基以它罕见形式出现时就可能和错误的碱基形成碱基对(图 21 - 2)。这种碱基化学结构形成的改变叫做**互变异构移位**(tautomeric shift)。

图 21 - 3 表示从 G - C 到 A - T 的转换突变可能是互变异构移位的结果。如果没有 DNA 聚合酶校正活性的话(见第 12 章),那么互变异构移位所产生的突变要比实际上发生的要多得多。

在 DNA 复制中少量碱基的增加和缺失也能自发发生(图 21 - 4),这可能由于新合成链或模板链错误的环出(跳格)而产生的。若是新合成的链的环出可增加一个碱基对;若模板链的环出则会缺失一个碱基对。DNA 中少量碱基的增加和减少,除增加或

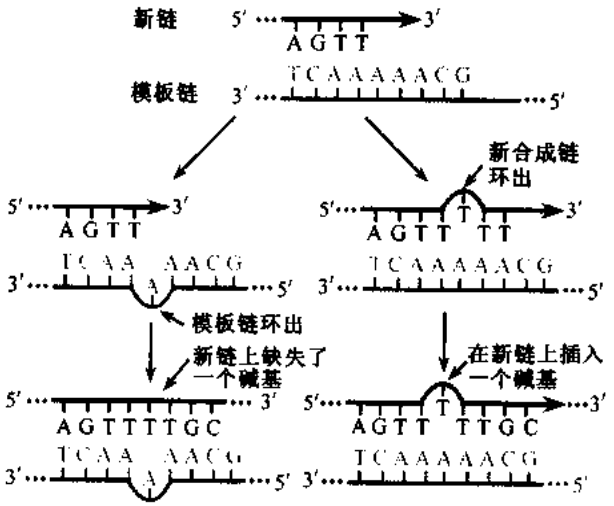


图 21 - 4　在 DNA 复制中由于 DNA 的错误环出自发产生了碱基的插入和缺失。
(参考 Peter J. Russell: Genetics, 3rd ed., 1992, Fig. 18 - 7)

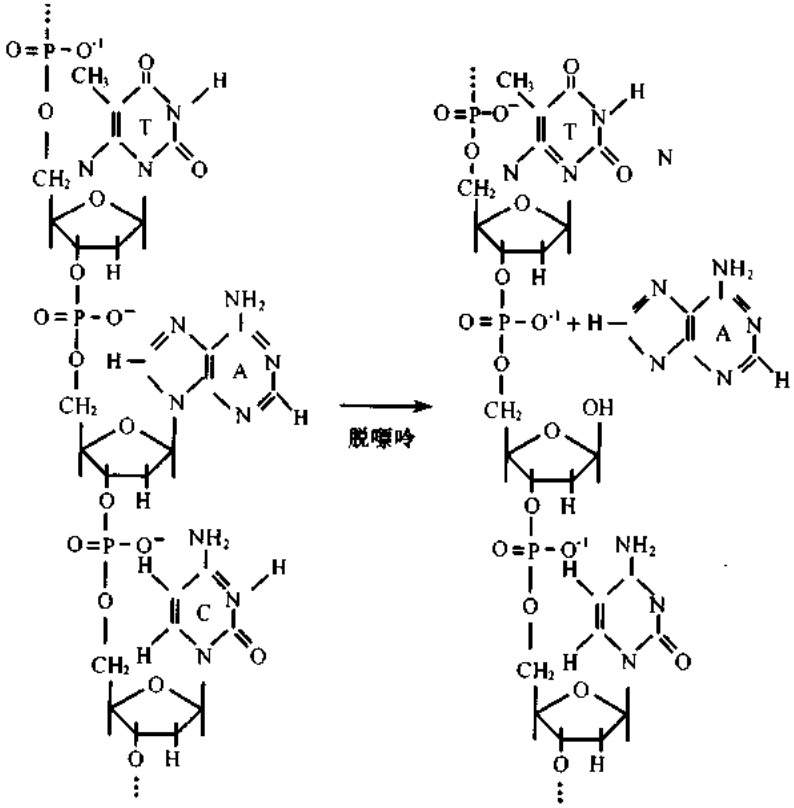


图 21 - 5　从 DNA 单链上脱嘌呤,磷酸二酯键未断裂。(仿 Peter J. Russell: Genetics, 3rd ed., Fig. 18 - 8)

减少 3 个碱基以外，都会引起移码突变。

2. 自发的化学变化

引起自发突变的两种最为常见的化学变化是特殊碱基**脱嘌呤**（depurination）和**脱氨**(基)(deamination)作用。在脱嘌呤时，脱氧核糖和嘌呤之间的糖苷键断裂，A 或 G 从 DNA 上被切下来(图 21－5)，在培养中的哺乳动物细胞增殖期有数以千计的嘌呤通过脱嘌呤作用而失去，若这种损伤得不到修复的话，在 DNA 复制时，就没有碱基特异地与之互补，而是随机地选择一个碱基插进去这样很可能产生一个与原来不同的碱基对，结果导致突变。

去氨(基)作用是在一个碱基上除掉氨基。例如在胞嘧啶上有一个易受影响的氨基，脱去这个氨基后产生了尿嘧啶(图 21－6)。在 DNA 中"U"并不是一个正常的碱基，修复系统就要除去大部分由 C 脱氨而产生的 U，使序列中发生的突变减少到最小程度。如果 U 不被修复的话，在 DNA 复制中它将和 A 配对，结果使原来的 C-G 对转变成一个 T-A 时，产生了碱基转换突变。后面我们将要谈到亚硝酸的脱氨作用是另一个脱氨引起突变的例子。原核和真核生物的 DNA 含有相对少量的修饰碱基——5－甲基胞嘧啶(5mC)。5mC 也可脱氨，但产生的是 T(图 21－7)。由于 T 是 DNA 中正常的核苷，所以没有一种修复机制能觉察和改正这种突变。

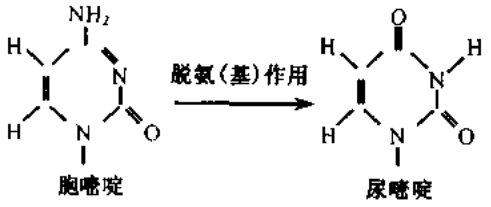


图 21－6 胞嘧啶脱氨基产生尿嘧啶。
(仿 Peter J. Russell: Genetics, 3 rd ed., 1992, Fig. 18－9)

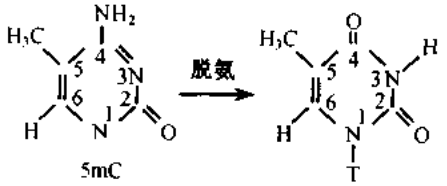


图 21－7 5mC 脱氨产生胸腺嘧淀。
(仿 Peter J. Russell: Genetics, 3rd ed., 1992, Fig. 18－10)

5mC 的脱氨的结果使 5mcC：G 转换或 T：A，且 5mC 脱氨突变却不能通过修复机制得到校正。基因组中 5mC 的位点常常是**突变热点**（mutational hot spots），即在此位点发生突变的频率要比别处高得多。

还有第 3 种自发突变是**氧化作用损伤碱基**(oxidatively damaged bases)。有活性的氧化剂，例如过氧化物原子团(O_2^-)，过氧化氢(H_2O_2)和羟基(－OH)等需氧代谢的副产物，它们可导致 DNA 的氧化损伤，导致突变和人类的疾病，胸苷氧化后产生胸苷乙二醇，G 氧化后产生 8-氧-7,8 二氢脱氧鸟嘌呤、8-氧鸟嘌呤(8-O-G)或"GO"，GO 可和 A 错配，导致 G→T。

二、诱发突变

自发突变的频率是很低的，诱变剂可以增加突变的频率。常用的两类诱变剂是**放射线**和**化学物质**，两者涉及特殊的作用机制。

1. 射线

用电滋波谱中比可见光的波长要短一些的部分（即少于 1μm）构成非离子射线和离子射线，前者以紫外线为例，后者包括 X 射线、γ 射线以及宇宙射线。由于当波长减小时能量增大，因此紫外线比可见光的能量要高一些；离子射线比非离子射线的能量大；γ 射线要比 X 射线能量大。离子射线能穿透组织，因此 X 射线可用于诊断。通过与原子碰撞并释放出电子，此电子再与其他原子碰撞释放更多的电子，使得高能复射也能产生离子。沿着每个高能射线的轨迹能发现一串离子，它们能起动很多的化学反应，其中包括突变。电离射线可诱导基因突变和染色体的断裂。

紫外线（ultraviolet light ray，UV）是非电离化的，那是由于它们没有足够能量来诱导离子化。但紫外线是常用的诱变剂，高能可使它杀死细胞。遗传学和医学界都广泛应用 UV 的这一性能，UV 常被用来杀菌。UV 还能引起突变，这是因为 DNA 中的嘌呤和密啶有很强的吸收光的能力，特别是对波长为 254～260nm UV。这种波长的 UV 能通过 DNA 光化学变化初步诱导基因突变，使 DNA 合成延伸衰减。UV 能量此 X 射线要低，其穿透力也十分有限。

UV 对 DNA 的作用之一是在同一条链中在两个相邻的嘌呤分子之间。或在双螺旋的两条链的嘌呤之间形成异常的化学键。大部分是 DNA 的两个相邻的 T 被诱导形成共价链，产生了胸苷二聚体，常以 $\widehat{TT}$ 表示（图 21－8）。

通常突变是由于 UV 的照射使 $\widehat{TT}$ 的持续存在或未能修复所致。胸苷二聚体可以

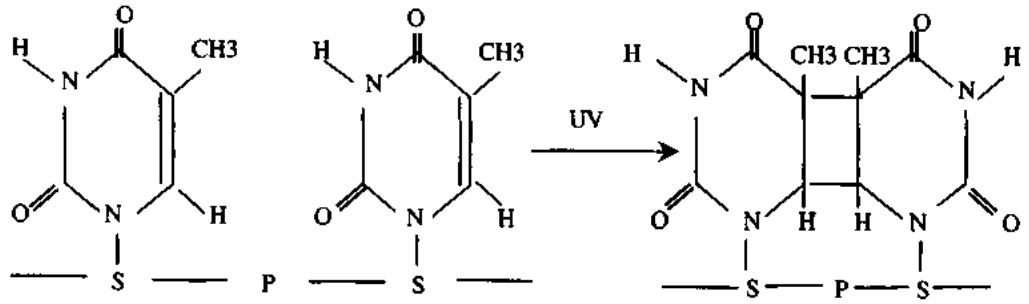


图 21－8　两个相邻的 T 经 UV 照射形成 $\widehat{TT}$。

被特殊的细胞系统所修复（见第 22 章）。产生这种修复需要的细菌中一种特殊的宿主酶系统，人们推测在真核细胞也存在这种系统。这种系统称为 SOS 修复系统，也称为**错误倾向修复**（error prone repair）。当细胞遭到大的损伤时，会被诱导产生一种反应来阻止细胞死亡。SOS 的作用不限于 UV 导致的损伤，其他种类的 DNA 损伤也能使此系统活化。

2. 化学诱变剂

（1）碱基类似物

碱基类似物是一类化学结构与 DNA 中正常碱基十分相似的化学制剂，例如 5－溴尿嘧啶（5-bromouracil，5-BU），它和 T 很相似，仅在第 5 个碳原子上由溴（Br）取代了 T 的甲基。5-BU 有两种异构体，一种是酮式，另一种是烯醇式，它们可分别与 A 及 G 配对结合，这样在 DNA 复制中一旦掺入 5BU 就会引起碱基的转换而产生突变（图 21－9）。

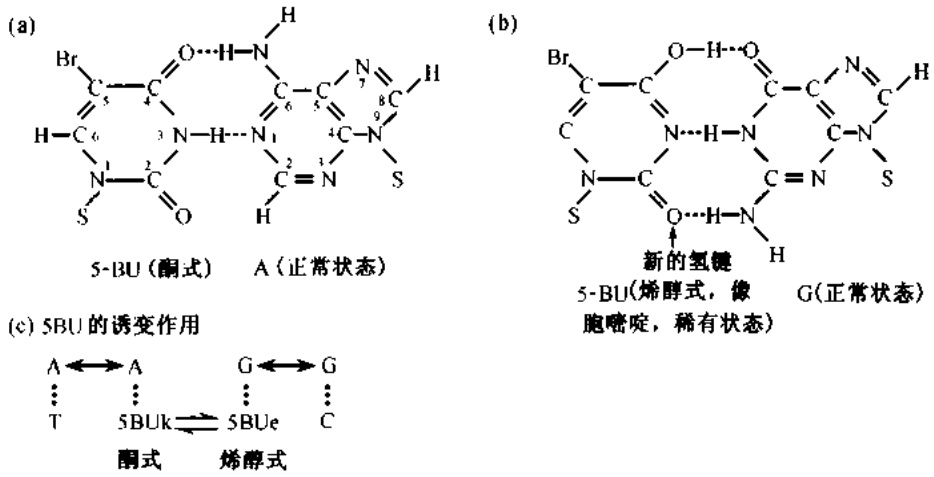


图 21－9 5BU 的酮式和烯醇式结构及与 A、G 配对，(a、b)，在 DNA 复制中 5BU 的掺入导致 AT→GC、GC→AT 转换突变(c)。(仿 Peter J. Russell: Genetics. 3rd ed., 1992, Fig. 18－13)

氨基嘌呤(2－aminopurine 2－AP)也是碱基的类似物，它也有两种异构体，一种是正常状态，另一种是稀有的状态，以亚胺的形式存在，它们可分别与 DNA 中正常的 T 和 C 配对结合(图 21－10)。当 2－AP 掺入到 DNA 复制中时，由于其异构体的变换而导致 A：T→C：G 或 G：C→A：T，其机制与 5－BU 相似(图 21－11)。

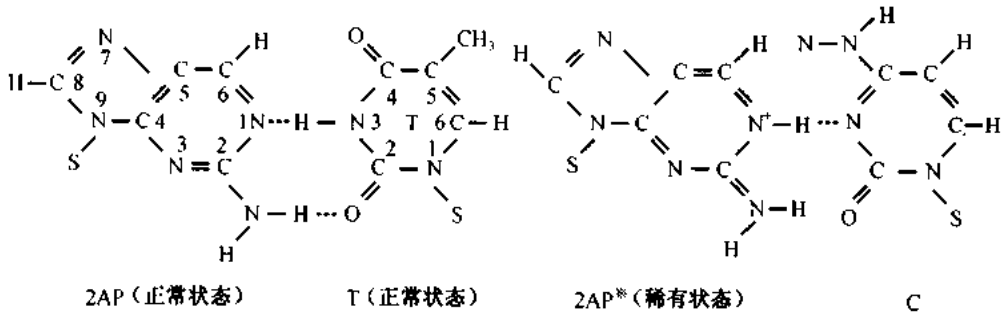


图 21－10 2AP 的两种异构形式及和 T、C 的结合。(仿 Peter J. Russell: Genetics, 3rd ed., 1992, Fig. 18－14)

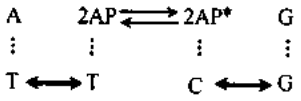


图 21－11 2－AP 诱变机制。

最后是用于治疗艾滋病（获得性免疫缺损综合征，acquired immunodeficiency syndrome，AIDS）的一种药物叫**叠氮胸苷**（**AZT**，azidothymidine），它也是一种碱基类似物。艾滋病毒，又称为人类免疫缺陷病毒－1(human immunodeficiency virus-1, HIV-1)，是一种反转录病毒，其遗传物是 RNA，但当病毒侵入细胞后通过反转录酶将基因组 RNA 反转录成一个 DNA 拷贝即 cDNA。这个 DNA 整合到宿主细胞的基因组 DNA 中，然后能进行一系列的新的蛋白合成，从而产生新的病毒。而 AZT 能作为 T 的类似物掺入 DNA 中。AZT 在病毒 RNA→DNA 的阶段是反转录酶的底物，但在细胞中它却不是 DNA 聚合酶的合适底物。这样 AZT 的作用是一种选择性的毒物，可以抑制病毒 cDNA 的产物，阻断新的病毒的合成。

(2) 碱基的修饰剂

有的诱变剂并不是掺入到 DNA 中，而是通过直接地修饰碱基的化学结构，改变其性质而导致诱变，如亚硝酸，羟胺，烷化剂等(图 21－12)。

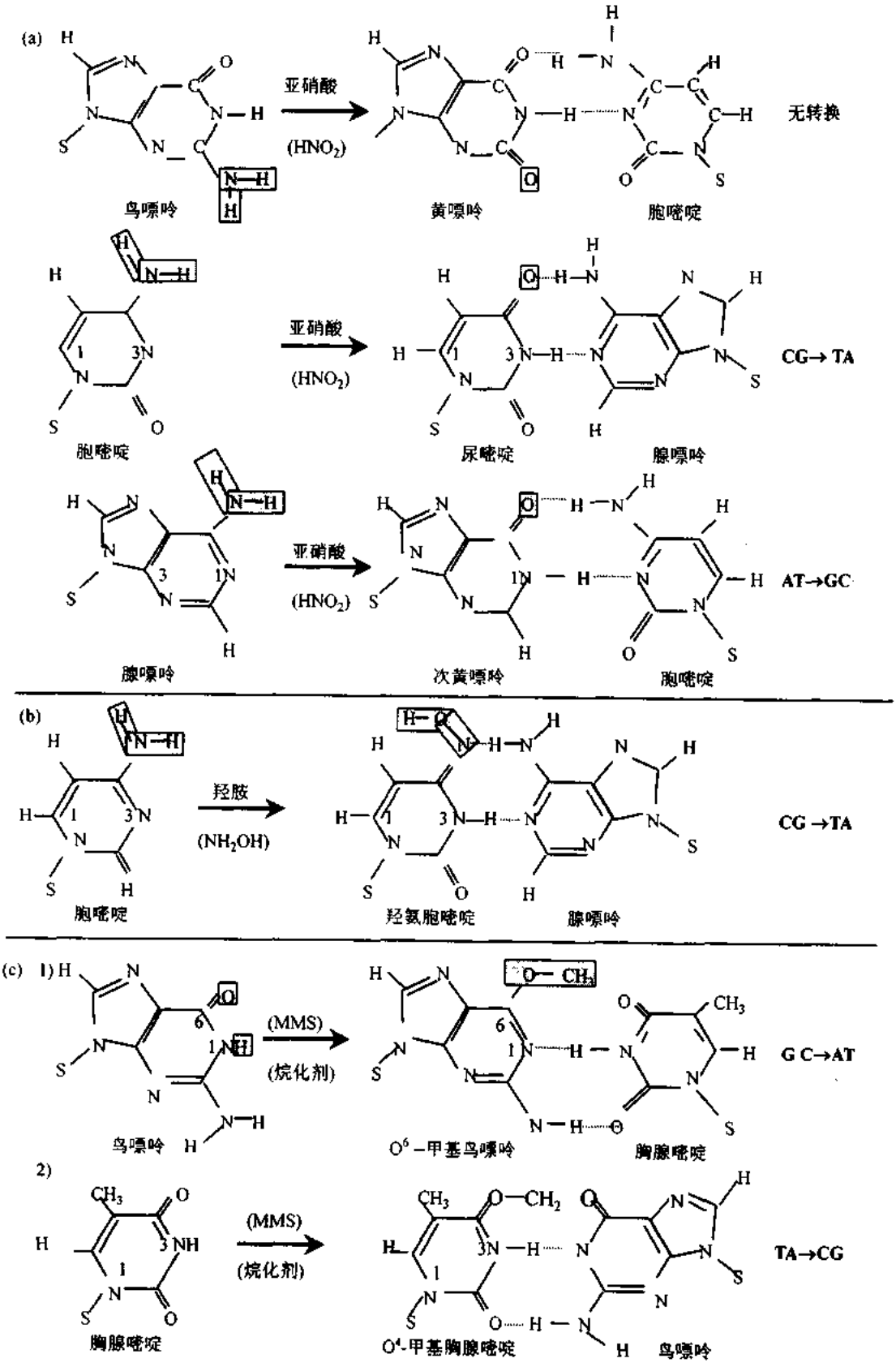
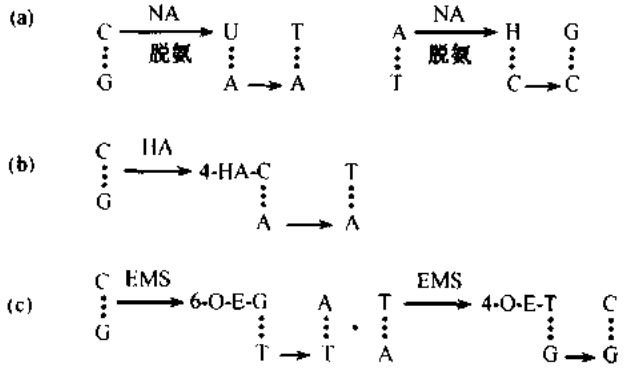


图 21－12　三种碱基修饰剂的作用。(a)亚硝酸修饰 G、C、A；(b)羟胺修饰 C；(c)甲基黄酸乙酯修饰 T。
(仿 Peter J. Russell: Genetics, 3rd ed., 1992)

亚硝酸（nitrous acid，NA）具有氧化脱氨的作用，可使 G 第 2 个碳原子上的氨脱去，产生黄嘌呤(xanthine,X)，次黄嘌呤（H）仍和 C 配对，故不产生转换突变。但 C 和 A 脱氨后分别产生 U 和次黄嘌呤(H)，产生了转换(图 21－12a)，使 C：G 转换成 A：T，A：T 转换成 G：C(图 21－13a)。



(a)C·G $\xrightarrow{NA}$ T·A; A·T $\xrightarrow{NA}$ G·C (b)C·G $\xrightarrow{NA}$ T·A

(c)G·C $\xrightarrow{EMS}$ A·T; T·A $\xrightarrow{EMS}$ C·G

图 21－13　三种碱基修饰剂导致碱基的转换。

另一种碱基修饰剂是羟胺(图 21－12b)，它只特异地和胞嘧啶起反应，在第 4 个 C 原子上加上－OH，产生 4－OH－C，此产物可以和 A 配对，使 C：G 转换成 T：A(图 21－13b)。

甲基黄酸乙脂(EMS)，氮芥(NM)，甲基黄酸甲脂(MMS)，亚硝基胍(NG)等都属于烷化剂，它们的作用是使碱基烷基化，EMS 使 G 的第 6 位烷化，使 T 的第 4 位上烷化，结果产生的 O－6－E－G 和 O－4－E－T 分别和 T、G 配对（图 21－12），导致原来的 G：C 对转换成 A：T 对；T：A 对转换成 C：G(图 21－13c)。

(3)DNA 插入剂

插入突变剂（intercalating mutagens）包括原黄素（proflavin），吖啶橙（acridine orange），溴化 3,8－二氨基－5－乙基－6－苯基菲啶翰(etnidium bramide)及 ICR 的复合物(图 21－14)等。它们通过插入到 DNA 双螺旋双链或单链的两个相邻的碱基之间，起到插入诱变的作用。它们的化学结构都是扁平的分子(图 21－14)易于在 DNA 复制时插入。

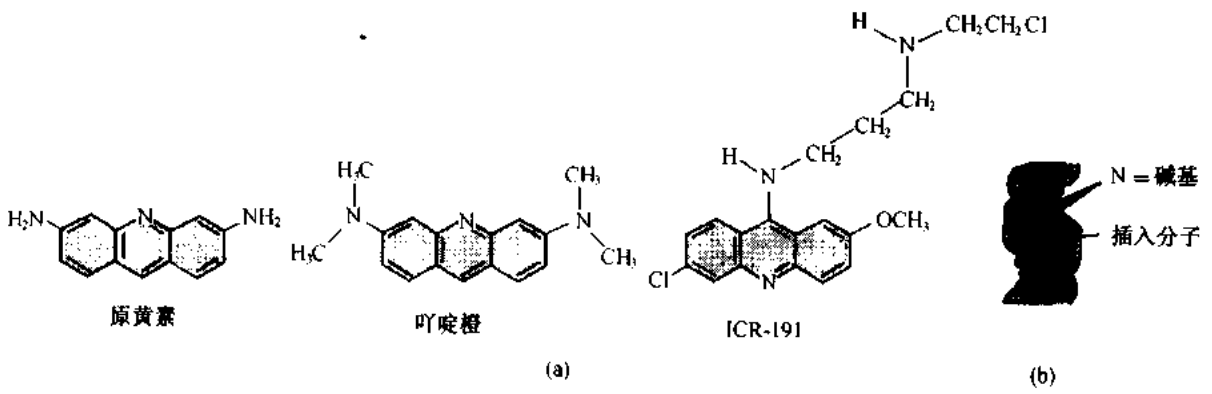


图 21－14　插入突变剂。(a)普通插入剂原黄素，吖啶橙和 ICR－191 的结构；(b)一个插入剂分子滑入堆积在 DNA 分子中间的两个氨基之间，此可能导致单个碱基对的插入或缺失。(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 18－20)

若插入剂插在 DNA 模板键两个相邻碱基中，合成时新合成链必须要有一个碱基插在插入剂相应的位置上，以填补空缺，这个碱基不存在配对的问题，所以是随机选择的。新合成链上一旦插入了一个碱基，那么下一轮复制必然会增加一个碱基（图 21－15a）。如果在合成新链时插入了一个分子的插入剂取代了相应的碱基，而在下一轮合成前此插入剂又丢失的话，那么下一轮复制将减少一个碱基（图 21－15b），这样使的新合成链增加或减少了一个碱基，引起了移码突变。

（4）DNA 的体外定点突变

无论是人工诱变还是自然突变都是随机的过程，在育种工作中人们常用诱变的方法，然后经筛选，试图获得理想的品种。事实上人工诱变仅增加突变的频率，而并不改变突变的方向，因此这种筛选在一定意义上仍是“靠天收”，碰运气，带有一定的盲目性，这是因为人类以前未能找到一种定点突变的方法。20 世纪 80 年代以后人们利用分子克隆技术建定了**体外定点突变**（site-specific in *vitro* mutation）的方法，并利用这种方法在已知的 DNA 序列中增删或转换核苷酸，特异地产生实验所设计的突变，现在这种方法已广泛地应用于蛋白质的结构与功能的研究，转录调控元件结构和功能的研究，以及构建新的载体或嵌合基因，改变限制性酶切位点，内含子的剪接位点等等。因此这项技术的建立者加拿大的 Michael Smith （1985）于 1993 年获得了诺贝尔化学奖。

体外定点突变的具体方法有三种：①聚核苷酸介导的用单链模板定点突变（图 21－16）；②双引物法定点突变；③用掺入 U 的单链为模板进行聚核苷酸介导的体外定点突变。以上三种方法虽不同，但原理都一致。现以第一种方法为例来介绍其基本原理。

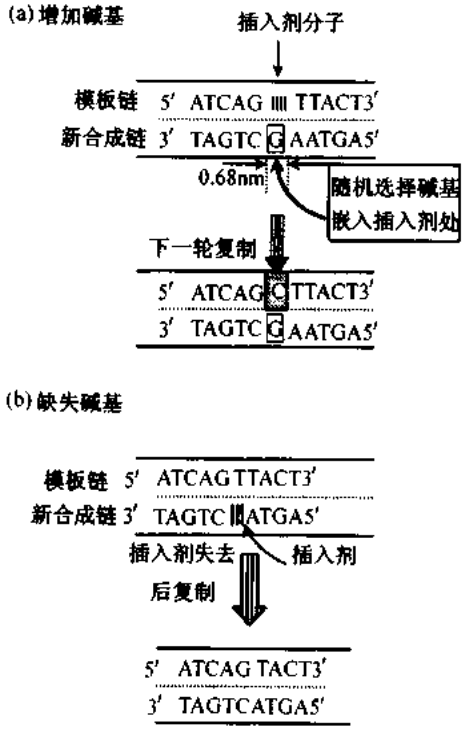


图 21－15　插入突变导制碱基的增加(a)和减少(b)。
（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 18－16）

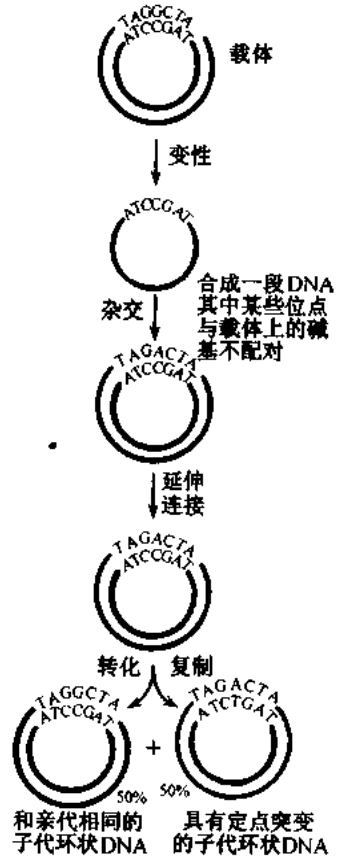


图 21－16　体外定点突变原理。

用已知序列的环状 DNA 变性后为模板，人工合成一段引物，我们所要设计的定点突变就在引物中，也就是说合成引物时不是完全和模板互补，而是在某个位点上有意识地让碱基“突变”，和模板上的碱基不能配对，由于其他的碱基是互补的，所以仍然通过复性，可以使引物和模板特异结合，然后在体外加 DNA 合成系统，让其延伸，最后的裂缺可用连接酶连接，然后再将这杂合双环 DNA 转化到细菌中。由于复制是半保留的，经克隆后将有一半的后代环状 DNA 产生了定点突变，另一半是和正常的亲代链一样。

用这种方法可以研究未知基因的功能，即将某基因进行定点突变，然后再导入细胞，观察表型有何变化，以探讨正常基因的功能。这些实验方法叫做**反向遗传学**(reverse genetics)。传统遗传学是从表型追溯到相关的基因，反向遗传学是以基因出发来研究其功能。

(5)Ames 检测法

到目前为止我们在遗传学研究中都使用各种诱变剂（放射线和化学制剂）来诱发突变。而在我们的环境中也可能存在着很多化学制剂具有诱变剂的作用。例如在动物实验室中约有一半的天然化学品(天然的化学品是指一些正常土产品，例如食物)是**致癌剂**(carcinogens)，它们可以引起癌的产生。动物实验中合成化学品中也有一半是致癌剂。(合成化学品包括杀虫药，药物和食品添加剂)。近来的资料表明有丝分裂剂(促使细胞分裂的化学品)具有明显的致癌作用。很多的致癌剂并不是诱变剂，但致癌剂却能引起有丝分裂。

各种实验方法被用于检测环境中是否存在可以致癌的化学品。有一种方法是用小鼠或大鼠的近亲繁殖系建立起来的化学致癌剂的检测方法。近亲繁殖系小鼠的遗传背景是相同的。经设立对照组，统计分析就可以提供有关致癌的信息。虽然这项实验比较昂贵，而且时间也较长，但仍被研究室列为常规。

早在 1970 年 Bruce Ames 发展一项快速的测验方法，现称为 **Ames 测验**（Ames test)。这种方法用沙门氏细菌(*Salmonella typhimutium*)作为实验生物。一般常用的是：组氨酸营养缺陷型品系。其中一个是 *his* 突变，可以通过取代一个碱基对而达到回复突变，而另一个品系的 *his* 突变要通过移码突变才能产生回复突变（这些品系也带有其他的突变但和本实验无关）

Ames 法的原理主要是依据两点：①化学物质引起的诱变往往不是直接的，它要经过生物的消化、吸收、特别是高等生物肝脏中的有关酶的作用以后再起作用。也它本来是可以直接起到诱变作用的，但经肝脏中的酶作用后失去了诱变能力；也许正好相反，本来它并不具有诱变的能力，但经过酶的作用后反而获得了诱变的能力。因此经肝脏中酶的作用才能较真实地反应在动物活体中某种化学品的实际的诱变能力。②诱变剂仅改变突变频率，而不影响突变方向，那么同样诱变剂也能导致回复突变。由于正突变是多方向的，因此必须采用反向选择的方向才便于研究，如用回突变只需用基本培养基进行筛选即可，步骤就要简单的多。

Ames 检测方法如图 21－17 所示，首先一组雄性大鼠进行腹腔注射芳香族化合物如多氯联苯油溶液等。以诱导大鼠肝脏酶系的活性，4 天后杀鼠取肝，捣碎后离心制备成肝酶(S9)的悬液，然后将待测物和 S9 以及沙门氏菌的突变株(his^-)(点突变或移码突变)混合后倒平板，若能产生回复突变的待测物则可判定为诱变物。在此实验中还有一组对照实验，即只加 S9 和沙门氏菌的突变品系，而不加待测物，若也有回复突变产生表明是自发的突变，可以作为对照来进行比较(图 21－15)。Ames 方法的特点是快速

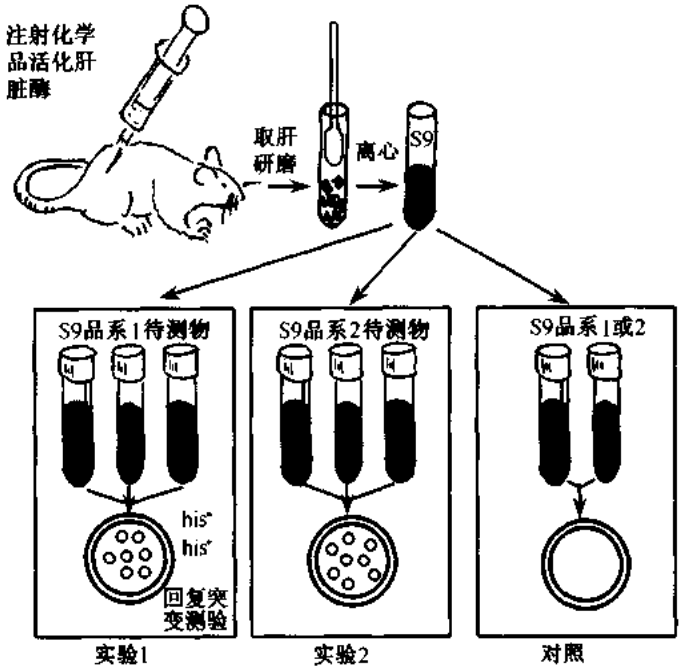


图 21－17　Ames 测验。$his^- \rightarrow his^+$ 回复突变实验。

简便，也较为准确。经对几百种物质进行测试表明约 90%致癌物具有诱变作用。

除了 Ames 方法以外，人们还常采用染色体畸变，姐妹染色单体变换(SCE)频率的改变以及微核技术来检测诱变物。这些方法都是观察待测物质的直接作用的结果，可作为初步筛选。

第三节　突变和人类疾病

一、自发突变和人类的疾病

1. 重复序列异常所引起的遗传病

DNA 序列分析提示出大量的人类遗传病是由突变所引起的。前面讨论的细菌突变也使我们了解突变导致人类疾病的机制。

很多人类的疾病是由于重复序列的缺失（deletion）或重复（duplication）所致（图 21－18）。例如线粒体的脑脊髓病(mitochondrial encephalomyopathies)就是一组影响到中

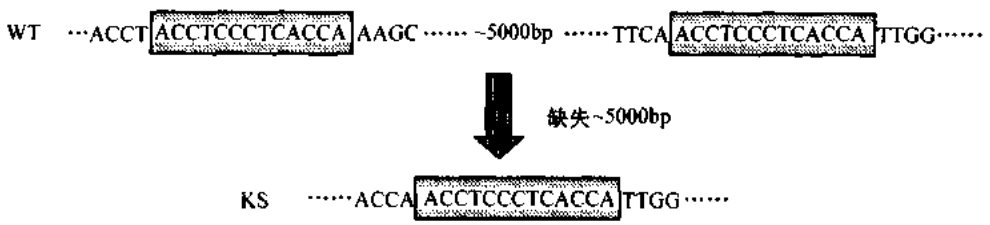


图 21－18　野生型(WT)DNA 序列和 Kearns-Sayre/ 慢性眼外肌瘫痪综合征(KS)患者的 DNA 缺失。
(仿 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 18－11)

枢神经系统或肌肉的疾病，其特征是氧化磷酸化紊乱（与线粒体的功能有关）。此是线粒体结构改变产生的，更确切地说是重复序列之间产生缺失所致。图 21－18 表示这种缺失。有趣的是在 *E. coli* 中也有类似的缺失。第二个例子是 Fabry 病，这是一种多糖神经鞘酯的先天性缺陷，是由 X 连锁编码 α－半乳糖苷酶基因突变所引起。这种突变中很多是基因重排，使得短的正向重复序列中产生缺失和重复。表 21－4 表示导致 Fabry 症的短重复序列中的重排断裂点。在珠蛋白基因中的发生的有些基因重排，可导致贫血和地中海贫血。所有的这些缺失都是通过滑动错配的机制即跳格（图 21－4）或重复序列之间的重组而产生的。

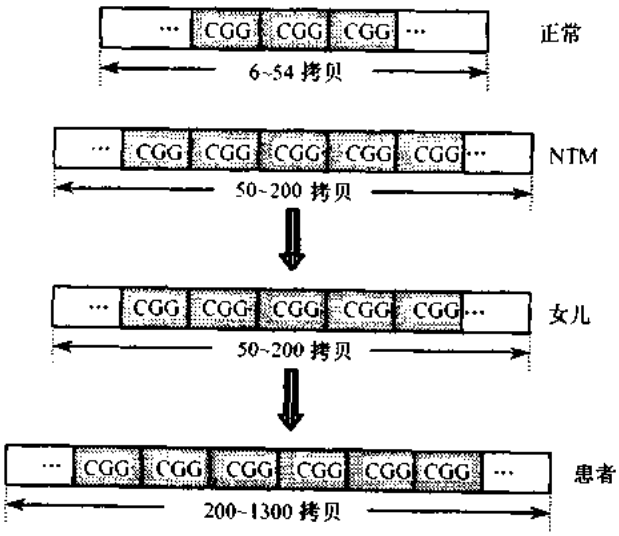


图 21－19　在脆性 X 综合征中 *FMR－1* 基因内 CGG3 个核苷酸扩增。*NTM*：正常男性。（*Griffiths A. J. F. et al.*：An Introduction to Genetic Analysis，5th ed.，1993，Fig. 18－12.）

很多的遗传病是由一种共同的机制产生的。这种机制是 3 对碱基重复序列的扩增（expansion of a three-base pair repeat），这种情况在脆性 X 染色体综合征（fragile X syndrome）中是存在的（图 21－19）。这种综合征男性发病率为 1/1500，并伴有大睾丸症。女性发病率为 1/2500。细胞学检查表明患者的 X 染色体长臂末端上有一个脆性位点，在体外可引起染色体断裂。

脆性 X 染色体综合征常由表型正常但带有一个脆性 X 染色体的男性传递的，遗传率约 20%，他把这个脆性 X 染色体传给女儿，女儿表型也正常。这种男性就称为**“正常”传递男性**（normally transmitting males，NTMs）。但 NTM 的外孙常出现此综合征。脆性 X 染色体综合征是由于 *FMR－1* 基因编码序列中的重复序列（CGG）发生突变。此病患者表现特殊的甲基化，此是在 CpG 基因簇附近的突变所引起的，结果使 *FMR－1* 基因表达变弱。

为什么在有些带有脆性 X 染色体的个体中会出现综合征，而另一些个体就不出现综合征呢？看来关键是 *FMR－1* 基因中 CGG 重复的数量。在正常人类中 *FMR－1* 基中 CGG 重复序列数量的变化范围是 6～54，通常带有 29 个重复序列，CGG 重复序列的变化和 *FMR－1* 基因编码蛋白中的精氨基酸数量的变化相一致（CGG 是 Arg 的密码子）。NTMs 和他们的女儿二者都有大量的重复序列。范围在 50～200。这种增加称为**前突变**（premutation）。所有前突变的等位基因都是不稳定的。带有综合征的男性和女性以

及很多的女性携带者都有附加 DNA 的插入。重复序列的量为 200～1300。显然，在 NTMs 和他们的女儿中所发现前突变等位基因里的重复序列的数目显示了一定的域值范围，超过这个范围显示突变表型，而未超过此范围仍是属正常的个体。

这种 CGG 重复的机制是在 DNA 合成中滑动错配所造成（图 21－4）。这种特高频

表 21－4　在哺乳动物胚期重排中和短的正向重复有关的断裂点序列

重排	5′断裂点	3′断裂点
Hb Leiden	TC CTGA GGAG	AGGAGAAGTC
Hb Lyon	GGGCAAGGTG	GGTGAACGTG
Hb Freiburg	TGAAGTTGGT	GTTGGTGGTG
Hb Nireroi	GGTTCTTTGAGTC	AGTCCTTTGGGGA
Hb Gun Hill	AGTGAG CTG CA CTG T	GACAAG CTGC ACGTC
Hb Tochigi	TATGGGCAAC	CTAAGGTGAA
Hb St. Antoine	TGATGGC CTGG	GC CTGG CTCAC
Hb Coventry	TAATGCC CTGG C	C CTGG CCCACAA
γδβ－地中海贫血 1	TC CCAG CACT	GAAAAGTCTG
Hb Bk	GGTAT CTGG AG	AATTTCTATTA
印度 HPFH	CGCGCCA CTG CAC	ATCCCCACTATAT
荷兰 β⁰－地中海贫血	AACCAAATTTGCAC	GAGAAAATTTTTGC
土耳其 β－地中海贫血	GTCTACCCTTGG	TTGGAC CCAG AG
土耳其 $(\alpha)^{20.5}$	CCTAGGCAACA	TAAGGGCCACG
RBI	AGCTTTTATACTTGA	TGAATTTAAACATAA
Pro-α2(I)	TTTCTTTCTAAG	GTGGTTTCCCTG
Fabry 家族 A	GAAC CCAG AAC	AGCTCCACCTC
Fabry 家族 B	CATCAAGGGTA	TTAAAAGACTT
Fabry 家族 D	AGCTTAGA CAG GTA	AATGTAGATAAAGA
Fabry 家族 E	CAG CAG AACTGGGG	GGAAAGAACTTTGA
Fabry 家族 J	TTTAA CCAG G	TAATACATTT
	CAAAATTTTA	ATGTATAGGC

此表显示每个基因重排 5′和 3′断裂点区，缺失是在正向重复（黑体）和 4 个 5′及 3′附近的核苷酸之间，此和重组有关。CTGG/CCAG 4 核苷酸和 CTG/CAG 3 核苷酸用方框表示，Hb 为珠蛋白。

R. Kornreich, D. F. Bishop and R. J. Desnick: Journal of Biological Chemistry, 1990, 2. 9319

率的突变表明：在人类细胞中，当多于 50 个重复序列时，复制机制不可能是完全“忠实”的。可能会使序列和数量都发生改变。

另两种遗传病：**X－连锁脊髓和延髓肌萎缩**，也称为 Kennedy’s 症，也是由于 3 个碱基（CAG）对重复序列的扩增而引起。Kennedy’s 症的特点是进行性肌无力或肌萎缩。它是由于编码雄性激素受体的基因发生突变所致。正常的个体在此基因中平均有 21 个 CAG 重复序列。患者达 40～52 个。

强直性肌营养不良（Myotonic dystrophy）是一种常染色体显性遗传病，多表现为成年人的肌营养不良。这是在人类中由于序列的扩增引起疾病的又一个例子。患者于青春期发病，四肢无力，行走困难，易摔倒。肌肉萎缩常自四肢远端开始，逐步发展到面肌和颞肌等。足下垂，腱反射减弱或消失。家族易感性在连续几代中呈现病症加重的现象。这是由于在转录本 3′端的 CTG 三联体出现进行性扩增所致。正常人平均有 5 个

CTG 拷贝，有中度症状的患者具有 50 个拷贝，而重症状病人具有 1000 以上的 CTG 的三联体。

2. 诱发突变和人类的肿瘤

环境中的诱变剂可导致人类的癌症。UV 和黄曲霉毒素 B1(aflatoxin B1,AFB1)长期以来被人们分别疑为可引起皮肤癌和肝癌。现在通过人类癌基因的 DNA 序列分析提供了有关的证据。这个基因命名为 *P53*，它是很多肿瘤抑制基因之一，它们可以编码一种蛋白来抑制肿瘤的形成，人类癌症患者的一个很大的特点是抑癌基因发生了突变。

肝癌在南非和东亚流行，在这个区域中 AFB1 的大量存在与肝癌的高发病率是相关的。经分析发现在南非和东亚的肝癌患者中 *P53* 基因发生了 G→T 颠换突变，此是 AFB1 诱发突变的一种印记，但在同样的地区中肺癌，结肠癌和乳房癌中都不存在这种情况。另一方面来自低 AFB1 区域的肝癌患者中 *P53* 的突变也不是由 G→T 颠换而产生。用 Ames 法来检测 AFB1 的结果表明不同品系的沙门氏菌的回复突变率和 AFB1 的处理剂量呈正相关。以上证据使我们相信 AFB1 诱导的突变引起了南非和东亚肝癌的高发。

P53 突变测序也将 UV 和人类的皮肤癌紧密地连结起来，人类磷片上皮细胞癌的分析表明具有 *P53* 基因突变，它们都涉及到双嘧啶位点，大部分是 C→T 的替换。这是 UV－诱发突变的特点。另外各种肿瘤中 *P53* 都是由 CC→TT 双碱基的改变所致，这种情况仅在 UV－诱发的突变中才存在。

在现代人类生活的环境中要接触很多的化学物质，如药物、食品防腐剂、化妆品、农药，用于工业的一些原料污染物等等。其中很多成分具有致癌和致突变作用。比如包括食物防腐剂 AF2，食物熏剂乙二溴化物，抗血吸虫病的药物羟胺硫蒽酮(hycanthone)，各种染发剂以及工业原料氯乙烯等。虽然其中有的已控制使用，但又有很多化学产品问世，常常并不能对每种产品都进行严格的检测，这样就造成了一种潜在的危害。为了避免受到诱变物的损伤，对待各种化学品的使用都应持谨慎的态度。

第四节　DNA 的修复机制

原核和真核生物对于 DNA 的损伤都有很多的修复系统，如 SOS 系统。所有这些系统都是用酶来进行修复。其中有的系统是直接改变突变损伤，而另一些则是先切除损伤，产生单链的裂缺，然后再合成新的 DNA，将裂缺修补好。我们可以将不同修复途经分为以下几类。

一、直接修复

1. 通过 DNA 聚合酶校正修复

前面已阐明原核 DNA 聚合酶都具有 3′→ 5′的外切酶活性(见第 12 章)，可对复制中错误掺入的碱基进行较正，使得 DNA 复制中实际的差错率大大减少。这种 3′→ 5′

外切酶活性是原核生物 DNA 聚合酶的特点，在真核生物中 DNA 聚合酶 δ 和 α 不同，具有 3′→ 5′外切酶活性。

在介绍原核 DNA 聚合酶结构已说明 DNA 聚合酶 3′→ 5′外切酶功能是 ε 亚基承担的，若编码这个亚基的基因发生突变，那么就失去校正的功能。

2. 光复活反应

直接修复损伤的另一个例子是对由 UV 诱发的胸苷二聚体的修复，这个系统叫做光复活（photoreactivation）或**光修复**（light repair）。修复发生在波长范围为 320～370nm（蓝光）的可见光中，二聚体可直接恢复成原来的样子。催化光复活反应的酶叫做**光裂合酶**（photolyase），它是由 *phr* 基因编码的。此酶可被可见光的光子所激活，结合在二聚体上，剪切二聚体的部分。由于在所有被研究过的生物中都发现了这种酶，因此它可能是一种普遍性的酶，它的功能是沿着双螺旋“清扫道路”，寻找由胸苷二聚体产生的凸出部分。$\widehat{TT}$ 很少，而光裂合酶在 $\widehat{TT}$ 二聚体较少时非常有效。若有 6～4 个二聚体就难以奏效。

烷基转移酶（alkyltransferases）也是一种和直接修复损伤有关的酶，它们可以切除掉 NG 和 EMS 加在 G 的 O－6 位上的烷基。这种酶还可以将 O－6 甲基上的甲基转移到蛋白质的 C 上。当转移后，酶就失去了活性，因此这种修复系统在烷基水平足够高时达到饱和。

二、切除修复

1. 一般切除修复

UV 光诱发 $\widehat{TT}$ 的第二种修复过程叫**切除修复**（excision repair），由于这过程并不依赖于光的存在故又称为**暗修复**（dark repair）。此修复机制是 1964 年由 R.R.Boyce 和 P.Howard-Flander 及 R.Setlow ，W.Carrier 两个组同时发现的。他们分离到某些对 UV－敏感的 *E.coli* 突变株，经 UV 照射后，它们在暗处具有比正常情况高的多的诱发突变率。这些突变体是 *Uvr*A（UVrepair）突变体。*Uvr*A 突变体只有在光照时才能修复二聚体，表明它们缺乏暗修复系统。野生型的生物在暗处能修复二聚体，野生型的结构以 *Uvr*A$^+$ 来表示。

E.coli 中的切除修复系统并不仅修复嘧啶二聚体，也能修复 DNA 双螺旋的其他损伤。图 21－20 表示这个系统的修复机制为先切后补。双螺旋的结构变形可被 UveABC 内切酶识别，这种多亚基的酶由 3 个基因（*UVA* 、*UVB* 和 *UVC*）编码。其功能如图 21－21 所示。

首先是 *Uvr*AB 组合在一起识别嘧啶二聚体和其他大的损伤。然后 *Uvr*A 解离（此需 ATP）而 *Uvr*C 和 *Uvr*B 结合，*Uvr*Bc 复合体在损伤位点的两侧面剪切，在损伤位点 5′端相距 7 个 nt 处，3′端相距 3～4nt 处剪切，此过程也需 ATP，*Uvr*D 是一种解旋酶（表 21－5），它可以帮助 DNA 解旋，让两个缺口间的单链片段（包含损伤位点）释放出来。大肠杆菌的一些酶也能在体外切除嘧啶二聚体。这种酶包括具 5′－3′外切酶活性的 DNA pol Ⅰ和单链特异的外切酶Ⅶ。在这些切除中，任何一个

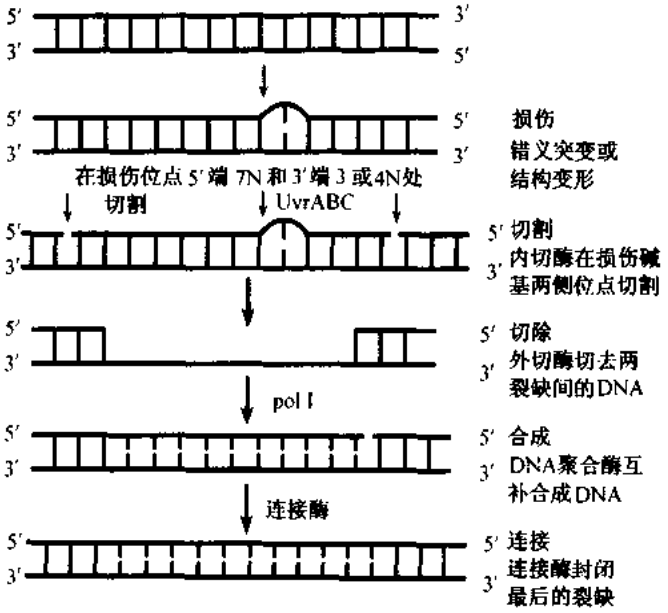


图 21－20　*E. coli* 切除修复系统。

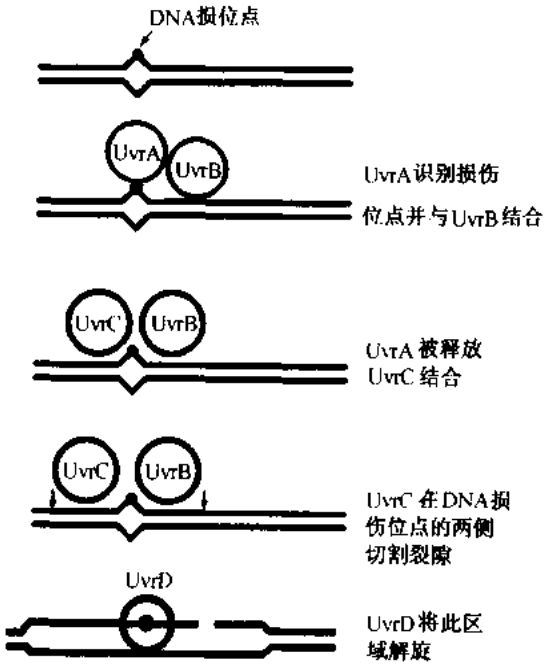


图 21－21　UvrABC 系统的作用，UvrAB 识别损伤 UvrBC 在 DNA 切割，UvrD 解旋。（参考 B. Lewin: Genes Ⅵ, 1997, Fig. 16－13）

发生突变 *E. coli* 都会失去切除嘧啶二聚体的能力。DNA pol Ⅰ好像是在体内担任最主要的切除任务。

被切除的 DNA 片段平均为 12nt，这种模型称为**短补丁修复**（short-patch repair），在这个修复合成中涉及的酶可能也是 DNA pol Ⅰ。

短补丁修复可解释 99% 的切除修

表 21－5　与 DNA 复制或修复有关的基因

基因	酶的活性
mutD = *dnaQ*	DNA pol Ⅲ
MutU = *UvrD*	*Uvr* 系统的 DNA 解旋酶
mutH, *U*	损伤修复系统的成分
mutL, *S*	同上，也切除 C-T 错配中的 T
mutY	切除 A-G 和 A-C 错配中的 A

摘自 B. Lewin: Genes V, 1994, Table 20.3

复。余下的 1% 和 DNA 延伸取代有关，长度可达约 1500nt，甚至可延伸到 9kb 以上，这就是**长补丁修复**(long - patch repair)。这个模型也需要 *Uur* 基因和 DNA pol Ⅰ。这两种模型的不同之处在于短补丁修复是由细菌细胞中特有的功能，而长 - 补丁修复必须要通过损伤来诱导。长补丁修复可能作用于复制叉附近区域的损伤修复。对二者的一些基因产物尚不清楚。

此修复系统可切除很多类型的损伤，包括 UV 光产生突变以及黄曲霉素(AFB1)结合产生的化学附加物以及苯并嘧啶的环氢化物等引起的损伤。其他生物中也具有此修复系统。在人类中这也是一个极其重要的修复途径。有一种罕见的遗传病叫**着色性干皮病**（xeroderma pigmentosum）就是一种切除修复酶的缺陷，本病患者的暴露部位易发生色素沉着，皮肤萎缩，角化过度和癌变，大部分患者在 30 岁前死于皮肤癌。由此可见切除修复系统途径的作用是如何重要。

2. 特殊切除修复途径

有些损伤太细微的以致产生的变形小到不能被 *Uur*ABC 系统所识别，因此还需要其他的切除修复途径。

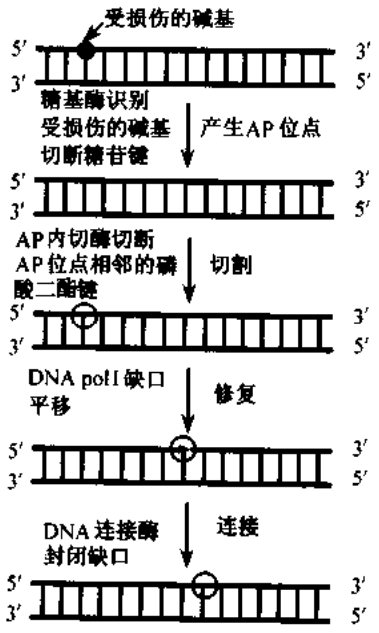


图 21 - 22　AP 位点的修复。

(1)AP 核酸内切酶修复途径

各种细胞中都有一种核酶内切酶，它附着在自发丢失了单个嘌呤或嘧啶的位点上，这个无嘌呤(apurinic)和无嘧啶(apyrimidinic)位点就叫做 **AP 位点**。AP 内切酶是细胞所必需的，因为自发地脱嘌呤作用经常发生，这种酶可以通过剪切 AP 位点的上的磷酸二酯键使链断裂。这是起始切除 - 修复；进一步由 3 种酶：外切酶，DNA pol Ⅰ和连接酶进一步进行修复（图 21 - 22）。

由于 AP 内切酶修复途径的功效，它可能是其他修复途径最后的一步。这样如果损伤的碱基对能被切除的话，那么就留下一个 AP 位点，AP 内切酶就能使野生型完全得到恢复。

(2)糖基酶修复途径

DNA 糖基化酶（glycosylase）并不能剪切磷酸二酯键，但可以剪切 N - 糖苷键。释放出改变了的碱基，产生一个 AP 位点(图 21 - 22)。这样再经过 AP 内切酶切割磷酸二酯键，再由 DNA pol Ⅰ的外切酶活性 5′ - 3′的合成活性进行修复，最后由连接酶连接。

DNA 糖基化酶有很多种，其中之一是尿苷 - DNA 糖基化酶，它可从 DNA 上切除尿苷，C 因自发脱氨基而产生 U，若不修复此可能导致 C→T 的转换。实际上在 DNA 中只有 A：T 配对而没有 A：U 配对，这是由于偶尔掺入的 U 可被识别和切除之故。若 U 是 DNA 中的正常成分，那么这种修复就不需要了。

还有一种糖基化酶，它可识别和切除由 A 脱氨而产生的次黄嘌呤(H)。另一些糖基化酶可切除被烷化的碱基(例如 3 - m - A，3 - m - G 和 7 - m - G)，开环嘌呤，氧化损伤的碱基以及在某些生物中的 UV 光化二聚体。一些新的糖基化酶仍不断被发现。

(3)GO 系统

mutM 和 *mutY* 基因产生的两种糖基酶一道作用，阻止突变产生的 8－O－G 或"GO"(图 21－23)。这些糖基酶形成了**GO 系统**。当 GO 丢失时 DNA 中会因自发氧化作用造成损伤，切除 GO 损伤的是由 *mutM* 基因编码的一种糖基酶。如果在复制时产生了氧化损伤形成 GO(8－O－G)，经复制 C 仍和 GO 配对，而 GO 和 A 配对，若得不到修复就导致了 C：G→A：T，但 *mutY* 编码的糖基本酶 MutY 可切除错配的 A(表 21－5)，经复制恢复为 GO：C(图 21－24a)。如果在复制的底物中掺入了 8－O－G，一种情况是它就可能和模链上的 A 配对，但细胞中有一种糖基酶 *MutT*，它起到 dUTPase 酶的作用，使 S－O－GTP→8－O－GMP；这样它就不能作为 DNA 合成的底物了。即使如此还是有部分 GO 逃过了 MutT 酶的"监视"，仍然掺入到 DNA 中和 A 结合，糖基酶 MutY 可以将错配的 A 切除，通过复制产生了 C：G。使 A：T→C：G。另一种情况是 GO 掺入到模板中，和 C 配对，当糖基酶切除 GO 后可排除产生 A：T 的危险，仍保持 C：G 对(图 21－24b)。

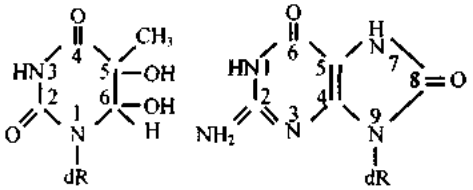


图 21－23 氧化作用产生的损伤碱基。
(仿 Anthony J. F. Griffiths *et al.*: An Introduction to Genetic Analysis, 4th ed., 1993, Fig. 18－10)

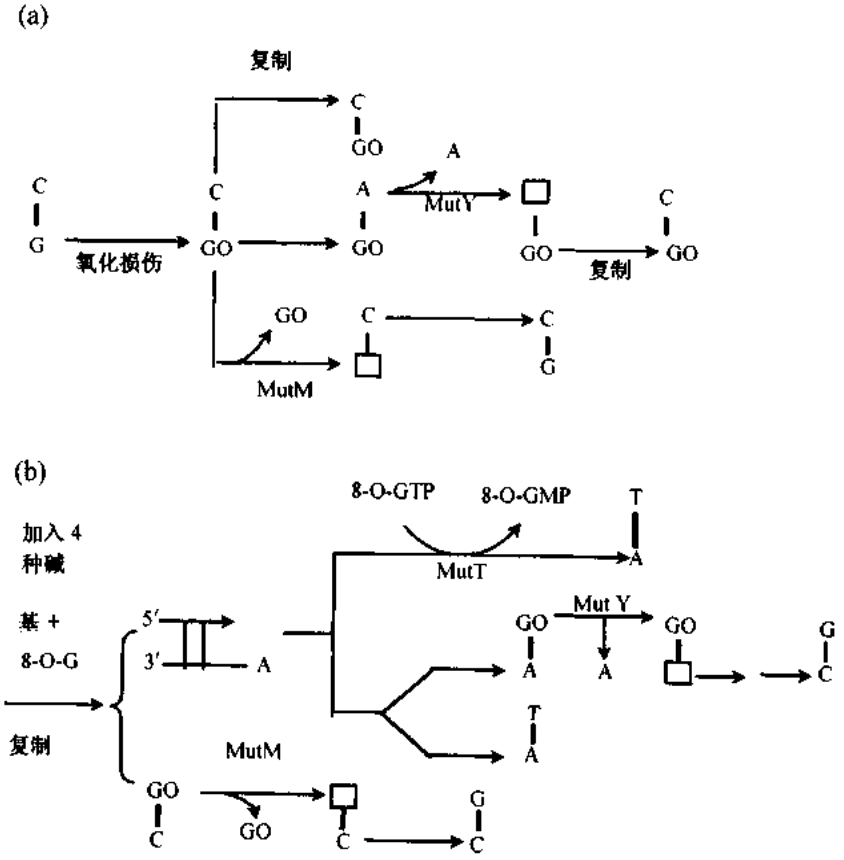


图 21－24 GO 修复系统。

三、复制后修复

1. 错配修复

有些修复途径能够识别 DNA 复制中出现的错配，这种系统叫做**错配修复**(mismatch repair)系统，假设要你设计一个可以修复复制错误的酶，那么这个酶应当是什么样的呢?看来至少它要具备以下三个功能:①识别错配的碱基对;②对错配的一对碱基要能准确区别哪一个是错的，哪一个是对的;③切除错误的碱基，并进行修复合成。

以上第二点是这个系统最为重要的特点，除非它能区分错误和正确的碱基，否则错配修复系统就无法决定切除哪一个碱基。例如 **5－m－C** 脱氨变成 **T**,有一个特殊的系统要将此修复成正常的序列。脱氨的结果使得 **C：G→G：T** 错配,这个系统必须将 **G：T** 修复成 **C：G**,而不是 **A：T**。

VSP(very short patch repair) 系统不能胜任此工作，损伤修复系统中的 *mutL* 和 mutS 也不能解决问题，它们是从错配的 G：T 和 C：T 中切除 T(表 21－5)。其他的如

表 21－6　在 *E. coli* 中和 DNA 损伤修复有关的基因产物及功能

基因	体内突变效应	基因产物	功能
UvrA	对 UV 敏感	内切酶的 ATP 亚基	起始切除 T 的损伤
UvrB	对 UV 敏感	内切酶的亚基	
UvrC	对 UV 敏感	内切酶的亚基	
UvrD	对 UV 敏感	DNA 解旋酶Ⅱ	DNA 解链
recA	1.重组缺陷 2.不能诱导修复损伤	40kDa RecA 蛋白	1.重组与修复所需的 DNA 链交换活性 2.起始 SOS 修复途径的蛋白酶活性
recB	重组缺陷	外切酶Ⅴ(ATP 酶)	重组和重组修复所需
recC	重组缺陷	外切酶Ⅴ(DNA 结合)	
recD	重组缺陷	外切酶Ⅴ(58k 亚基)	
sbcB	抑制 *recBC* 突变	外切酶Ⅰ	不知
recF	重组修复缺陷;在 *recBC*、*sbcB* 突变体中重组缺陷	不知	不知
recj		单链 DNA 外切酶	不知
recNOR	同 *recF*	不知	*recF* 途径的一部分
recQ	同 *recF*	DNA 解旋酶;ATP 酶活性	*recF* 途径的一部分
recE	重组缺陷	外切酶Ⅷ	不知
dam	UV 敏感,突变率增加	对 GATC 作用的甲基化酶	鉴别 DNA 的模板链
mutH	UV 敏感,突变率增加	外切酶	修复的成分
mutL	UV 敏感,突变率增加	不知	作用新链的系统
mutS	UV 敏感,突变率增加	识别错配	DNA 链的合成
UvrD	UV 敏感,突变率增加	DNA 解旋酶Ⅱ	DNA 解链
mutY	增加突变频率	腺嘌呤糖基酶	从错配中切除 A
ada	对烷化剂敏感	鸟嘌呤甲基转移酶	从 G 中切除 CH_3
alkA	对烷化剂敏感	甲基腺嘌呤糖化酶	切除甲基-A/甲基-G
umuC	减少倾向差错修复的效率	不知	不知
lon	UV 敏感,膜中隔缺失	结合 DNA 的 ATP 依赖性蛋白酶	可能是控制夹膜多糖的基因
lexA	调节 SOS 反应	阻遏物(22kDa)	控制很多基因

摘自 B. Lewin: Genes Ⅴ. 1994. Table 20.2

mutY 可以从错配的 C：A 中切除 A(图 21－22,表 21－5)。这些系统的功能都是直接地从错配碱基对中切除其中一个特定的碱基。*mutY* 的产物是一种腺嘌呤糖基酶,它可以产生一个 AP 位点来进行切除修复。同样不能确定哪一个是错误的碱基。

在 *E.coli* 复制中发生错配时，可以区分原来的模板链和新合成链，因为甲基化的 DNA 刚复制后只有在原来的亲本链上带有甲基，新合成的链尚未甲基化，因此两条链的甲基化状态是不同的（图 21－25），这就给复制差错的校正系统提供了一个基础。

dam 基因编码了一个甲基化酶,它的靶位点是 $\begin{matrix}GATC\\CTAG\end{matrix}$ 上的 A(表 21－5)使 A 成为 6－m－A，此半甲基化位点被用来作为复制中母链的标志，同样用于与复制相关的修复系统。

图 21－25 表示两种可能的错配修复模型，MutS 能识别错配位点，MutL 能作用新合成的链，*UvrD* 是解旋酶可使错配区双链打开，然后 SSB 结合在单链上防止复性，MutH 作为外切酶将含有错配碱基的新链片段切除掉,再由 DNA pol Ⅰ进行修补,最后由连接酶把裂缺封闭好。这样就完成了错配修复。

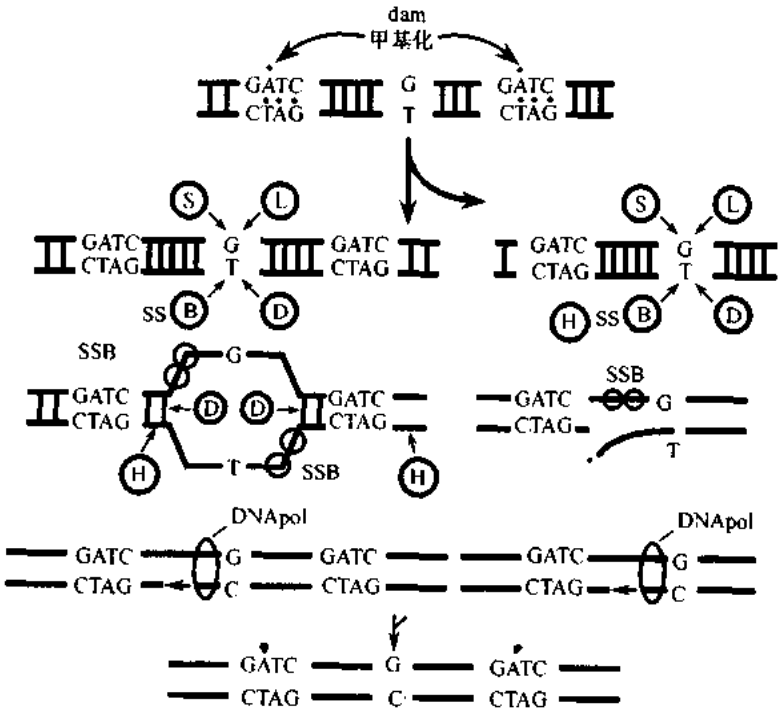


图 21－25 DNA 复制中的错配修复。（参考 A.J.F.Griffiths *et al*.: An Introduction to Genetics Analysis,4h ed.,1993,Fig.18－35. L:MutL;S:Mut S;H:Mut H;D:*Uvr*D,略修改）

2. 重组修复(recombination repair)

这个系统是 DNA 复制时模板链上含有损伤的碱基导致子链产生裂缺，所此修复系统也属于复制后修复。

若 DNA 上的一条链含有结构变形，如嘧啶二聚体，当 DNA 复制时二聚体就使损

伤位点失去做模板的作用，复制就跳越过这一位点。DNA pol 可能继续前进或者在嘧啶二聚体附近再重新开始合成。这样在新合成的链上留下了一个裂缺，使两个子链的性质不同。一条子链的亲代链上含有损伤，新合成的相应位点上有一个裂缺。另一条子链的亲代链是完好的，没有损伤，新合成的互补链也是正常的(图 21－26a)。恢复系统就利用了这条正常的子链。

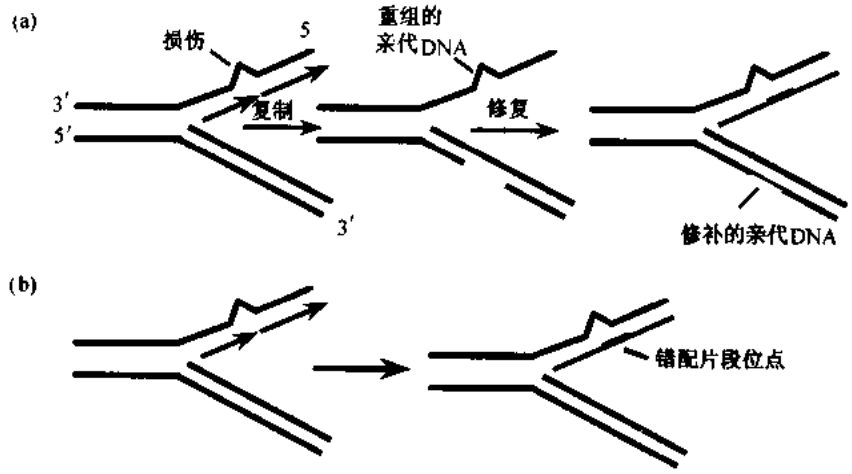


图 21－26　(a) 重组修复 (b) SOS 修复。(仿 A. Kornberg and T. Baker: DNA Replication 2nd ed., 1992, W. H. Freeman and Co., New York)

损伤子链的缺口由正常子链上的同源片段通过重组来填补，随着单链交换(single-strand exchange)受体的双链有一条是亲代有损伤的链，另一条经重组后变成为野生型的链；而供体双链有一条正常的新链和一条带有裂缺的亲代链，这个裂缺可通过一般的修复系统(DNA pol Ⅰ)来修复，最终产生一条正常的双链。这样损伤就限制在原来的链上，不至影响到新合成的 DNA。

在一个具有切除修复缺陷的 *E. coli* 中，*recA* 基因的突变使其失去所有的修复能力，人们试图在 *Uvr*⁻/*recA*⁻ 的细胞中进行复制产生一段 DNA 片段，预期其长度是在两个 TT 之间。实验结果表明根本得不到复制的 DNA，细胞由于缺乏了 *RecA* 的功能，妨碍了复制而致死。它解释了为什么 突变体(*Uvr*⁻/*RecA*⁻)的基因组中不能容忍超过 1～2 个二聚体，而野生型细胞允许存在多达 50 个。

recA 的突变体在遗传重组和修复中几乎完全得到鉴别。RecA 蛋白的功能是促使 DNA 链之间的交换。在 DNA 重组和涉及重组修复的单链交换中 RecA 都起着重要的作用。它还涉及差错倾向修复。

双突变体(*Uvr*⁻/*recA*⁻)的特点表明 *recA* 参与两种 Rec 途径，将单个基因突变体的表型和双突变体的进行比较，来测定二者的相关功能是部分相同还是不同。如果这些基因是在相同的途径中，那么双突变体的表型将和单个突变体的表型相同。如果这些基因是在不同的途径中，那么双突变体应影响到两种途径而不是一种，因此影响的表型应比单个突变体的要多。通过这种方法发现，一个 rec 途径和 *recBC* 基因有关，另一个途径涉及 *recF*(表 21－5)。

recBC 这两个基因编码外切酶 γ 的两个亚基，它们的活性受到此途径的其他成分

的限制。*recF* 的功能尚不清楚，其他的 *rec* 座位也都通过影响到重组和重组修复的突变而被鉴别出来。

Uvr 系统是负责切除大量的TT，而 Rec 系统是负责清除那些未被切除的二聚体。这些遗漏的二聚体虽数量不多，但常常是致命的。

3. SOS 修复系统

SOS 修复顾名思义是一种急救性的修复，忙中就难免有错，故也叫**差错倾向修复**(error prone repair)这种修复是为了保命也就管不了修补的片段是否正确。

Jeam Weigle 等曾用紫外线照射 λ 噬菌体然后再去感染细菌，而细菌又分为两组，一组是事先也用 UV 照射过，另一组是没有照射过，结果前一组中 λ 噬菌体的存活率反而高于后一组，这是什么缘故呢?

以上的现象称为 **UV 复活**(UV-reactivation)，也叫做 W 复活(Weigle 的第一个字母)，现在称为 **SOS 反应** (SOS response)。原来这是一系列相关蛋白与 RecA 蛋白及 LexA 阻遏物相互作用的结果。

碱基的损伤可能由于 UV 的照射，或者是由于交叉连结以及烷化作用等。而复制的抑制可能由各种因素而引起，包括 T 的失去，药物的加入或者一系列 *dna* 基因的突变等。SOS 反应增加了 DNA 损伤的修复能力，这是通过诱导长补丁修复系统和 Rec 重组修复途径这两个系统来完成的。另外细胞分裂的抑制，原噬菌体的溶源化都可被诱导。

SOS 反应的起始是通过损伤处理，RecA 的活化而引起。我们现在尚不知道损伤与 RecA 活性改变之间的关系。由于损伤引起的变化诱导 SOS 反应 RecA 被激活的，这也是和其他修复途径的不同之处。其他修复途径的酶是已存在于细胞中的，而 SOS 修复的酶系统是经损伤诱导才产生的。诱导的信号可能由 DNA 释放出的小分子组成的，或者是 DNA 本身某些结构变化。在体外 RecA 的激活需要单链 DNA 和 ATP 的存在。这样激活信号可能存在于损伤位点的单链区。无论信号是怎样产生的，它和 *RecA* 的相互作用是很快的，SOS 反应在产生损伤的几分钟内就发生了。

recA 的激活导致 *lexA* 的基因产物的剪切。lexA 是一种小分子蛋白(22kDa)，在被处理的细胞中相对稳定，它的功能是很多操纵子的阻遏物。此剪切反应是很特殊的；LexA 有一种潜在的蛋白酶活性，它可被 RecA 激活，当 RecA 被激活后又可导致 LexA 自我催化本身的裂解，这样 LexA 的阻遏功能失去了活性，并且同时诱导了它所结合的操纵子(图 21－27)。

LexA 靶基因很多都有修复功能。通过构建与 *lacZ* 基因的融合操纵子筛选出一系列基因，当用一种诱导损伤的药剂处理细胞后，β－半乳糖苷酶的产量就会增加，现在至少已鉴别出 5 个和 SOS 反应有关的基因，它们被称为 *din*(damage inducible)，它们包括 *recA*, *lexA*, *UvrA*, *UvrB*, *umuC* 和 *himA*。

某些 SOS 基因仅在细胞处理后才有活性，另一些在未处理的细胞中也有活性，但表达的水平随着 *lexA* 被剪切而增高。*UvrB* 是切除修复系统的一个成分，这个基因有 2 个起动子，一个功能依赖于 LexA，另一个易受它的控制。这样 *lexA* 被剪切后这个基因就用第二个启动子表达，剪切前就用第一个启动子表达。

LexA 阻遏靶基因时和它的 20bp DNA 结合，此结构区称 **SOS box**。其序列呈两侧对称，每个靶座位上都存在一个拷贝。在不同座位上的 SOS 盒也是不同的，但都有 8bp 的保守序列。像其他的操纵位点位点一样，SOS box 的操纵位点和启动子有重叠。在

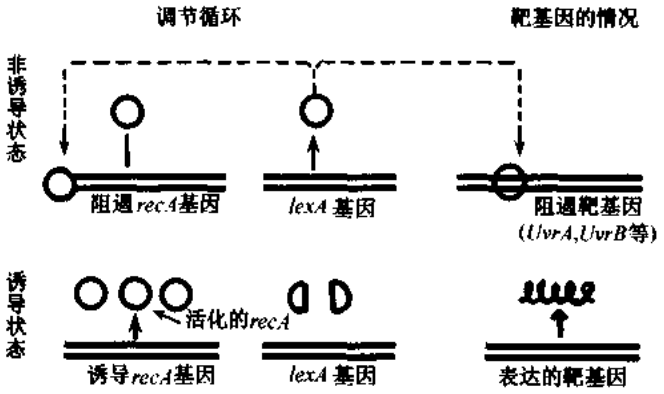


图 21－27　RecA 的诱导和调节。(仿 B.Lewin:Genes Ⅵ,1997,Fig.16－16)

lexA 座位属于自我阻遏物,那儿有两个相邻的 SOS box。

recA 和 lexA 在 SOS 调节循环中互为靶位点和靶蛋白。RecA 触发 *lexA* 的剪切,lexA 又是 *recA* 的阻遏物。SOS 反应导致了 RecA 蛋白和 lexA 阻遏的扩增。

recA 蛋白表达的增加对重组修复途径来说是必要的。据推导,*recA* 的水平从基本水平～1200 分子/细胞,提高约 50 倍,在被诱导细胞中这样高水平的 *recA* 足以使所有的 lexA 蛋白被剪切,从而抑制 lexA 对靶基因的阻遏作用。

这种调控循环的重要性在于使细胞迅速恢复到正常状态。当诱导信号去除后,RecA 蛋白失去使 LexA 去稳定的能力。此时,*lexA* 基因高水平表达;活化的 RecA 消失了,LexA 蛋白以非剪切形式迅速积聚并关闭 SOS 基因群。

RecA 也触发细胞中其他靶物质的剪切。*RecA* 激活时 umuD 蛋白被剪切;这种剪切激活 umuD 和差错倾向修复系统。

RecA 的激活也导致阻遏蛋白的剪切,包括各种原噬菌体的阻遏蛋白,如 λ 的阻遏蛋白这就可以解释为什么 λ 可以通过 UV 照射而被诱导裂解。溶源化的阻遏物 CI 蛋白被剪切,释放出的噬菌体进入裂解周期。

这些反应不是细胞的 SOS 反应,而是由原噬菌体来识别。对噬菌体而言生存的最好的方法是进入裂解周期,产生后代噬菌体。从这个意义上来说,原噬菌体的诱导是通过对相同的指示器(RecA 的激活)的反应使噬菌体从载体染色体上解离下来进入细胞系统。

所有已知的 RecA 的靶蛋白都是在多肽链的 Ala－Gly 二肽序列上剪切的。二肽两侧的任一侧必须是同源的,这表明蛋白质的一级结构是识别的重要特点。RecA 的两种活性是相对独立的。*recA* 441 突变使得在未进行诱导处理的情况下也能发生 SOS 反应,这可能是由于 RecA 保留了自发的激活状态。另一些突变可消除被激活的能力。两种突变都影响到 RecA 和 DNA 的结合能力。RecA 是 DNA 重组中重要的蛋白,后面将进一步加以介绍(见第 23 章)。

第二十二章

染色体畸变

基本原理

染色体畸变指染色体的结构或数目发生了异常的变化。染色体畸变可能是自发的，也可通过化学物质或放射线处理而诱发。

染色体畸变可能涉到部分染色体的结构改变，主要有四种类型：缺失（丢失了某个 DNA 片段），重复（增加了某个 DNA 片段），倒位（DNA 没有增加和减少，而是某一 DNA 片段的排列方向发生了改变）和易位（某一 DNA 片段从染色体的某一位置移到另一个新的位置上）。

一对同源染色体存在相同的结构改变称结构纯合体，若仅其中一条染色体结构发生改变叫结构杂合体。

染色体畸变会引起染色体的重排，可能影响到某些基因的表达。

染色体的畸变在人类中常产生畸形和异常，有些肿瘤的产生也和畸变有关。

重复、倒位和易位在物种的进化中常起着促进作用。

染色体数的改变分为两类，一类是整倍体，另一类是非整倍体。整倍体是增加或减少整套的染色体，非整倍体是增加或减少一条或几条染色体。

多倍体的植物较常见，其表型比二倍体粗壮肥硕。

多倍体分为同源多倍体和异源多倍体，同源多倍体部分可育，异源多倍体一般可育。但无论是同源三倍体还是异源三倍体都不可育，这些植物依靠营养繁殖来维持种的延续。

在动物中有些昆虫、蠕虫、蚌类、鱼类、两栖和爬行动物中都发现有多倍体。多倍体常依赖孤雌生殖。

非整倍体中分为缺体、单体、三体和多体。在动物中非整倍体都表现出畸形或异常。在人类中三体为常见的一种染色体病，21－三体称唐氏综合征。

在体细胞中也会出现非整倍体，人类的一些慢性髓性白血病患者的体细胞中常发现三体。

染色体畸变在动植物的育种中都已得到广泛的应用。

植物中的异源多倍体是新物种产生的一种重要的途径。

染色体的**畸变**（chromosome aberration）又称为**染色体突变**，包括染色体结构和数目的改变。染色体结构改变导致了染色体的重排。染色体数目的改变包括整套染色体的增减和单条或多条染色体的增减。基因突变在显微镜下观察不到，而染色体的畸变在显微镜下是可以看到并加以区分的。

很多染色体畸变导致细胞和生物功能的异常。这有两个基本反应，首先是染色体突变由于异常的基因数或位点而产生异常效应；第二，染色体畸变涉及到染色体断裂，这断裂常发生在基因的中部，从而破坏了它的功能。

染色畸变在生物学的各种不同水平上都是重要的。首先是它们提供了一种特殊的基因重排途径，通过研究染色体畸变能了解某些生物学问题。其次染色体畸变有重要的应用价值，特别是在医学及动、植物育种上的应用更为广泛。最后染色体畸变是生物进化过程的一个步骤。

研究正常或异常染色体组（chromosome set）及它们的特点的学科称为细胞遗传学，这是细胞学和遗传学相互融合的一门学科，这门学科以正常基因组（genome）作为标准或野生型，细胞遗传学家必须精通基因组的特点及界标。

第一节　断裂愈合和交换学说

各种染色体结构的改变都必须涉及某种方式的连接或染色体受到损伤和产生断裂。染色体的结构虽然也经有了各种模型，但仍还存在很多问题有待解决，这样也就限制了对染色体断裂的亚显微结构变化的解释。BrlnKly 和 Hittlman(1975)对人类染色体畸变的亚显微结构所提出的假设陷于进退两难的窘境。他们提出染色体断裂或交换(exchange)的机制仍未得到证实。染色体结构的改变一般是由于染色体的断裂和染色体片段的愈合（reunion）而产生的。染色体的断裂产生了损伤的（injured）末端，它和带有端粒的正常末端不同，具有粘性，易和其他粘性染色体末端相连接。

染色体断裂－愈合学说（breakage-reunion hypothesis）是由 Stadler （1931、1932)，Sax(1938、1941)，Muller(1932、1938、1954) Wolff (1961)及 Evans (1962) 相继建立。根据这个学说认为断裂是自发产生或者由诱变剂诱发，一般在原来的位置按原来的方向通过修复而重聚，此现象就称为重建（restitution)。若经重建改变原来的结构则称为**非重建性愈合**；若染色体断裂后的片段（不含着位点）丢失，留下了游离的断裂端，则称为**不愈合**。

染色体结构变异可分为四种类型，即①缺失(deletion 或 deficiency)：失去了部分染色体片段；②重复：(duplication)增加部分染色体片段；③倒位(inversion)：染色体片段作180°的颠倒再重接在染色体上；④易位(translocation)：两条非同源染色体之间产生部分片段的交换。

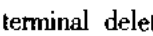
一对同源染色体其中一条是正常的而另一条发生了结构变异，含有这类染色体的个体或细胞称为**结构杂合体**(structral heterozygote)。若一对同源染色体产生了相同的结构变异，那么就称为**结构纯合体**(structral homozygote)。

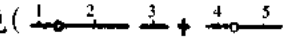
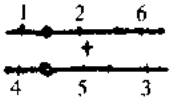
染色体结构变异能导致 4 种遗传效应：①**染色体重排**（chromosomal rearrangements)：染色体上遗传信息的顺序排列和邻接关系会发生改变。这会影响到基因的活性和表达，从而使表型发生一定的变异和异常。②**核型的改变**：若产生结构纯合体就会影响到核型的改变，这是进化中产生新物种的一种途径。③形成新的连锁群，由于染色体易位会产生新的连锁关系，改变了原来的连锁群，这也是物种进化的一种途径，但对个体来说常常产生表型异常。④减少或增加染色体上的遗传物质：染色体的缺失和重复会使染色体上遗传物质增加或减少，有的也影响到整个基因组的 DNA 含量，对于个体来说会产生表型变异或畸形，但对群体也是一种进化的重要途径，就整个的进化趋势而言，从低等到高等，从简单到复杂，基因组含量从少到多，这就涉及到染色体的重复。

第二节 染色体结构的改变

一、缺失(deletion)

1. 缺失的类型

(1)**末端缺失**(terminal deletion):染色体的末端发生缺失()

由于丢失了端粒,故一般很不稳定,比较少见,常和其他染色体断裂片段相重新愈合形成双着丝点染色()或易位();也有可能自身首尾相连,形成**环状染色体**(ring chromosome)。

双着丝粒染色体在有丝分裂中都可形成**断裂融合桥** (breakage fusion bridge) 或称**染色体桥**(chromosomal bridge),由于分裂时桥的断裂点不稳定,所以又可造成新的重复和缺失(图 22-1)

(2) 中间缺失 (interstitial deletion):染色体中部缺失了一个片段。这种缺失是较为稳定的,故较常见。

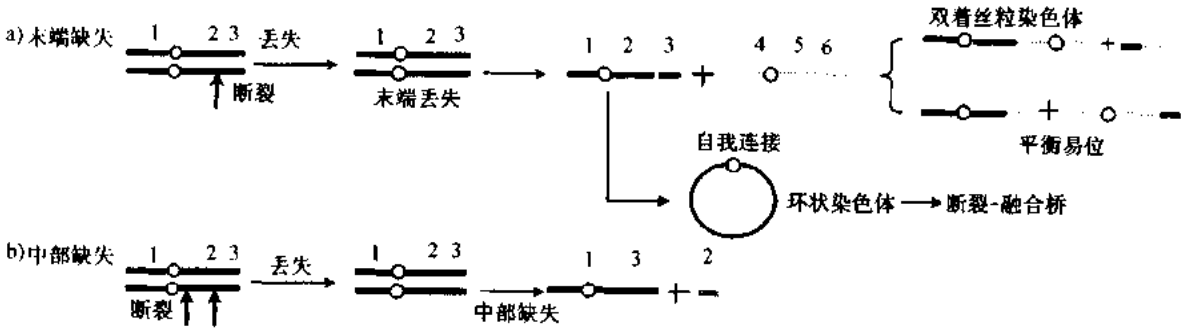


图 22-1 染色体缺失的类型及异常染色体的形成。

2. 缺失的产生

缺失的产生可能有以下几种原因:

(1) 染色体损伤后产生断裂(末端缺失)或非重建性愈合(中间缺失)(图 22-1)会直接产生缺失或形成环状染色体,经断裂融合桥而产生缺失和重复(图 22-8)。

(2) 染色体纽结:染色体发生纽结时若在纽结处产生断裂和非重建愈合就可能形成中间缺失(图 22-2,10)。

(3) 不等交换:在联会时略有参差的一对同源染色体之间发生**不等交换** (unequal crossing over) 结果产生了重复和缺失 (图 22-3),最早是在果蝇的棒眼突变型中发现的。

(4)转座因子可以引起染色体的缺失和倒位(见第 23 章)。

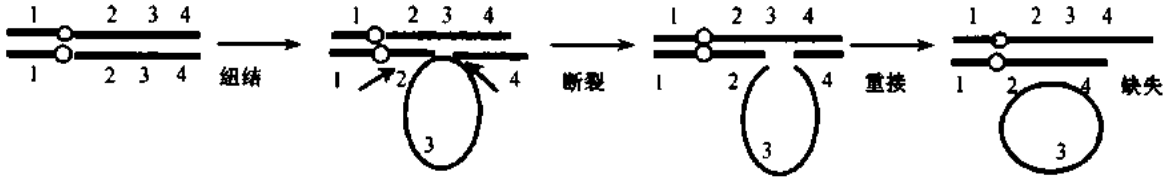


图 22－2　染色体纽结产生中间缺失。

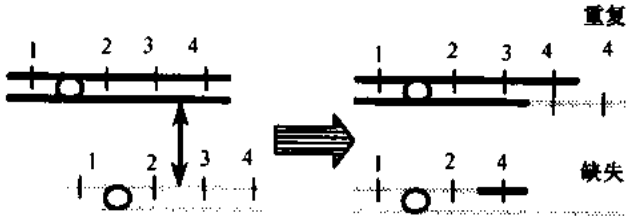


图 22－3　不等交换产生重复和缺失。

3. 缺失的遗传效应

缺失会产生以下几种效应。

(1)致死或出现异常

缺失的后果是严重的，常导致致死，但具体情况要看缺失片段的大小；是一对同源染色体同时都缺失了，还是只有一条缺失；以及缺失片段上所携带基因的重要性而定。发生缺失的配子常不育，尤其是花粉，而胚囊(8n)的耐受性略强，所以缺失多数通过卵细胞遗传的。缺失的纯合体比杂合体更难存活。

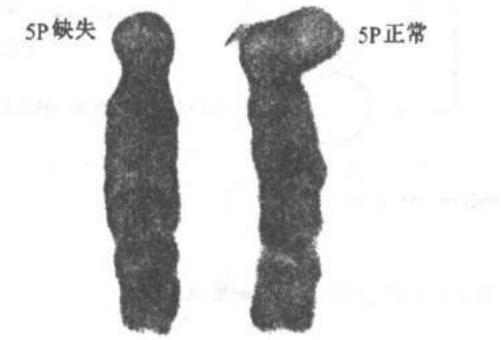


图 22－4　人类的猫叫综合征是由于第 5 号染色体短臂缺失所致。(引自 Russell, P. J.: Genetics, 1992, Fig. 21－24)

缺失小片段染色体有时虽不致死，但也会产生严重的异常，最有名的例子是人类的**猫叫综合征**(cri-du-chat syndrome，或 cat cry syndrome)。此综合征在 1963 年由 Lejeune 等首先报道了三例，患儿的哭声似猫叫，故得名。它是由于 5 号染色体短臂缺失($5p^-$)所致，(图 22－4)因而又称 $5p^-$ 综合征，发病率约为 1/5 万，女性多于男性，并有生长和智能发育不全。

(2)拟显性(pseudo-dominant)

显性基因的缺失使同源染色体上隐性非致死等位基因的效应得以显现，这种现象称为**拟显性**。例如玉米糊粉层的花斑有的是由于拟显性形成的(图 22－5)。玉米的胚乳是 3n，胚乳的颜色是由多对基因控制，只要其中任何一对发生突变，都不能形成色素，*C* 基因就是其中之一，它位于第 9 号染色体上。当胚乳细胞为杂合状态时(其他控制色素的基因都为野生型)仍产生色素，但是若含有"*C*"基因的片段缺失后，其相应

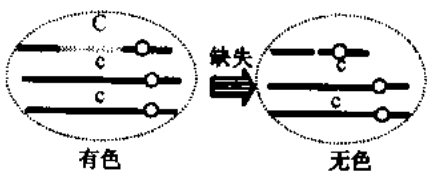


图 22－5　玉米胚乳颜色的拟显性。

的隐性基因的效应就得以显现，因此该细胞就不能形成色素。胚乳中有的细胞是正常的(产生色素)，有的存在 C 的缺失，因此胚乳出现了花斑。

另一个典型的例子是果蝇的**缺刻翅**(notch)(图 22－6)，即在果蝇翅的边缘有缺刻，胸部小刚毛分布错乱。当将红眼缺刻翅雌蝇和白眼正常翅雄蝇杂交所产生的 F_1 代，雌雄之比为 2∶1。

这是由于亲代雌果蝇的一条 X 染色体 C 区的 2－11 区域缺失了，而这一区域正是控制红眼和白眼的座位，因此到了 F_2 代雌果蝇中有 1/2 是白眼，是隐性基因 w^- 的等位座位发生缺失，使隐性基因得以表达的结果。在雄果蝇中有一半的 F_1 代其性染色体是由一条带有缺失的 X 染色体和 Y 染色体构成，Y 染色体没有相应的等位基因，不能利用互补效应来弥补缺失的功能因而致死(图 22－7)缺失的区域除了含有控制眼色的基因外还有控制翅形及刚毛分布的基因，因此同时出现了缺刻翅和胸部小刚毛分布错乱的表型。

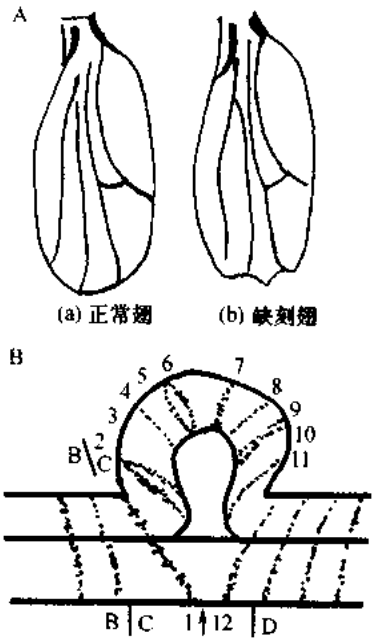


图 22－6 果蝇的缺刻翅和 C 区 2－11 的缺失。

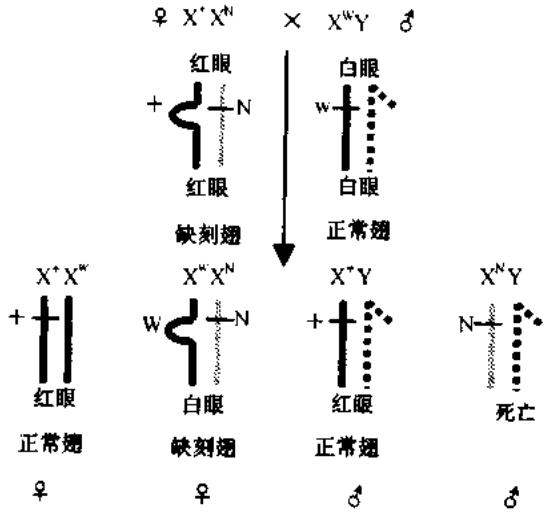


图 22－7 缺失引起白眼基因的拟显性表达。

4. 缺失与进化的关系

整个进化的趋势是基因组 C 值的增加，而缺失只能减小 C 值，因此它在进化中没有直接的作用。但还阳参属(*Crepis*)植物($2n = 6$)是惟一的例外，有人认为在千万年以前此属的染色体多于 6 条，后来由于着丝点和异染色质发生了缺失，常染色质易位到其他染色体上去了，于是造成了染色体数目的减少。故缺失在进化中的作用可能是从核中排出不必要的染色质减少染色体数。在进化中其他类型的染色体变异也会涉及到缺失，如易位在进化中起着重要的作用，但整臂易位或罗伯逊易位时就会涉及到着丝

粒或小片段染色体的缺失。

5. 缺失的应用

缺失常用来作为一种研究手段以此探测某些调控元件和蛋白质的结合位点，如 *E.coli* 的复制起始区的分析，以及酵母静息子的查明等。

二、重复

1. 重复的类型

(1) 顺接重复(tandem duplication)：重复顺序所携带的遗传信息的顺序和方向和染色体上原有的相同(1 2 3 2 3 4)。

(2) 反接重复(reverse duplication)：与上述相反，重复顺序所携带 DNA 顺序和原有的相反(1 2 3 3 2 4)。

(3) 同臂重复：重复的片段在同一条染色体臂上。

(4) 异臂重复：重复的片段在不同的臂上(1 2 3 1 4)。

(5) 异位重复(displaced duplication)：重复的片段来自其他非同源染色体(1 2 3 8 4)。

2. 重复的产生

(1) 断裂 - 融合桥的形成

由于断裂而丢失了端粒的染色体可自我连接形成环状染色体，复制后若姐妹染色单体之间发生交换则在有丝分裂后期可以形成染色体桥。附着在纺锤丝上的着丝粒仍向两极移动导致桥的断裂，常常导致了染色体的重复和缺失(图 22 - 8)。

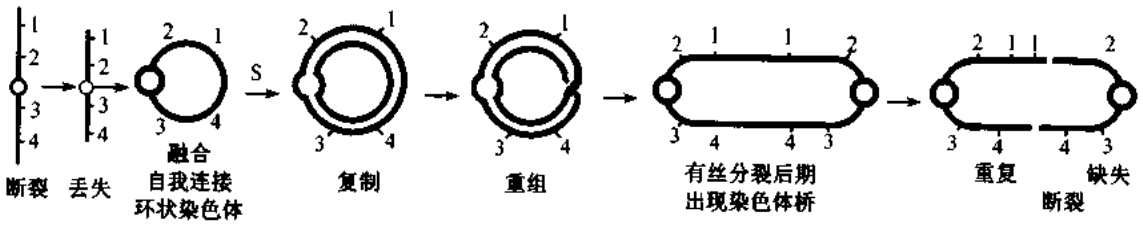


图 22 - 8　断裂融合桥的形成会产生染色体的重复和缺失

(2) 染色体纽结

一对同源染色体中的一条若发生纽结和断裂可能会产生反接重复和缺失(图 22 - 9)。

(3) 不等交换

不等交换如图 22 - 3 所示，最好的证据是人类珠蛋白结构异常的研究。前面已介绍人类血红蛋白由 2 条 α 链和 β 链构成，在 11 号染色体上有 β - 珠蛋白的基因家族。珠蛋白 ν - δ - β 连成一组。而有些地中海贫血的患者其 β 珠蛋白的亚基中只有部分

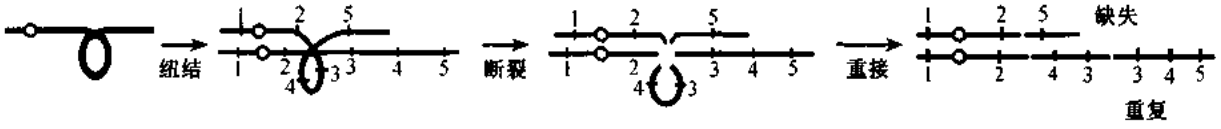


图 22-9 染色体纽结导致缺失和反接重复。

δ 和部分 β 组成（Lepore 珠蛋白）或者只有部分的 γ 与部分的 β 组成（Kenya 珠蛋白），这些异常的血红蛋白可能是不等交换产生的。Hb Lepore 是不等交换产生缺失的产物，必然还存在另一个相应的重复的产物，这就是反 Lepore（图 22-10）。Hb Kenya 发生的机制和 Hb Lepore 产生机制相同（图 22-10），因此目前尚未发现反 -Kenya（anti-Kenya），但将来总会发现的。以上例子有力地证明不等交换可导致染色体的缺失和重复。

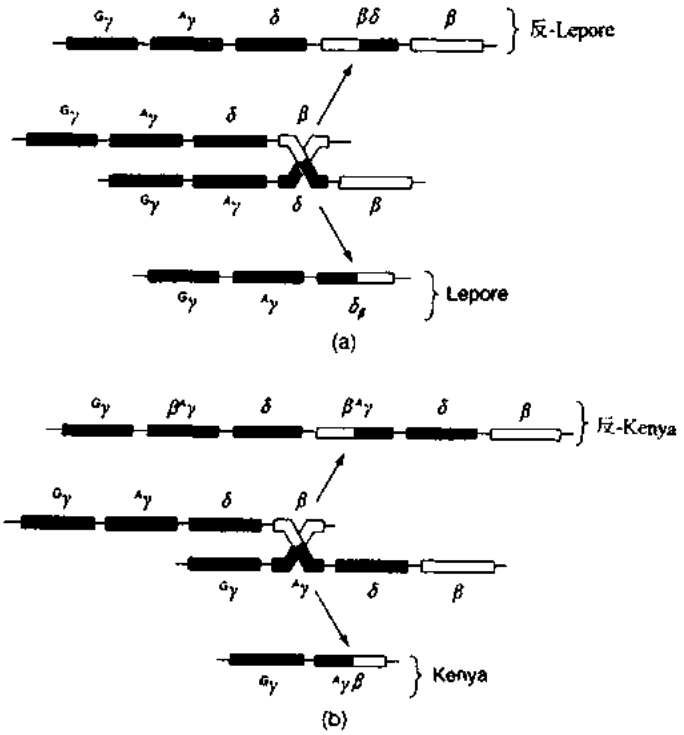


图 22-10 通过在 γ-δ-β 基因区域内的不等交换可产生人类特殊的血红蛋白异常。（转引自 Griffiths A.J.F. *et al*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 8-17）

(4)非同源性重组

如 λ 噬菌体的特异性转导和转座都可以将一个 DNA 片段导入受体基因组的靶位点中，从而产生重复。植物的转座因子 *Ac*, *Ds* 系统转座时，*Ds* 插入到新的位点，使得染色体增加了一个 DNA 片段（见第 24 章）。

(5) 易位的产物

易位的结果使某一条染色体增加了一个片段，而导致了移位重复。

3. 遗传效应

(1)**位置效应**（position effect）。一个基因随着染色体畸变而改变它和邻近基因的位

置关系，从而改变了表型效应的现象称位置效应。可能是随着位置的改变也改变了和5′端调控元件的关系和距离，从而影响了基因的表达。

从表型效应的稳定性来看，可把位置效应分为两类，一类叫做**稳定位置效应**(stable position effect)，又称为**S型位置效应**。如果蝇的棒眼(Bar)基因B位于16区A段，在正常情况下没有发生重复，果蝇的复眼由779个小眼组成。呈椭圆形，但16A区发生重复时复眼中的小眼数随之减少，复眼呈长形，故称为棒眼。16A重复的次数和排裂方式有规律地影响着棒眼的程度(图22－11，表22.1)。棒眼基因的重复引起的位置效应是稳定的，属于S型位置效应。

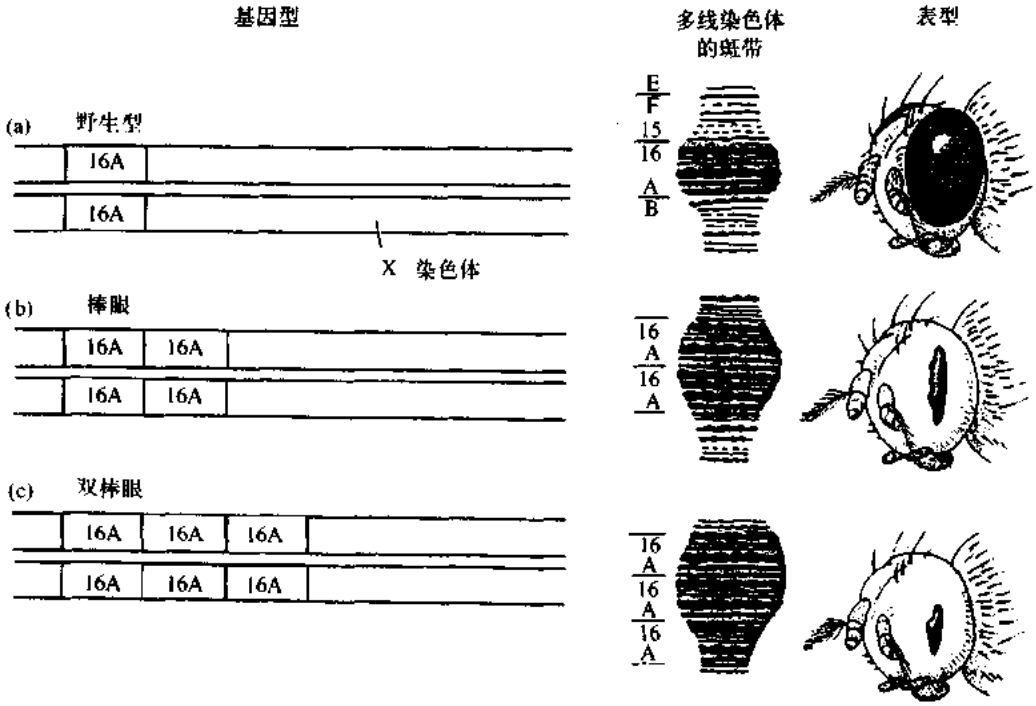


图22－11　果蝇基因的重复所引起的位置效应。图表明在唾腺染色体中，X染色体16A区的重复和棒眼大小的关系。(引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 19－6)

表22－1　果蝇棒眼基因的位置效应

B重复的情况	*B/B*	*BB/B*	*BB/BB*	*BBB/B*	*BBB/BB*	*BBB/BBB*
小眼数	779	358	68	45	36	25

另一种位置效应是不稳定的，称为**花斑位置效应**(variegated position effect)，也称为**V型位置效应**，常由易位而引起。

(2)表型异常。如果在细胞或个体中既有重复又有缺失，总的基因组平衡的话，除染色体重排引起的效应外，个体表型是正常的。但就整个的基因组而言若总量因重复而增加，那么某些基因及其产物的剂量也随之增加，如果这些基因或产物十分重要的话必然会引起表型异常。如一些不平衡易位的患儿，由于染色体的重复常引起智力低下或表型畸形。

4. 重复和进化

重复是进化中一种重要的途径，生物从简单到复杂的进化最根本的是基因组

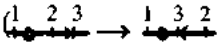
DNA 含量的增加和新基因的产生，而重复是增加基因组含量和新基因的惟一途径。比较有名的例子是珠蛋白的进化和乳酸脱氢酶(LDHA)同工酶的进化(见第 25 章)。

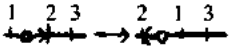
(1)通过重复可以研究位置效应。在细胞学研究中可以通过重复来给某一染色体进行标记，如 1931 年 C.Stern 和 H.Creighton，B.McClintock 两组所做的证实染色体交换实验都利用移位重复片段作为标记(见第 5 章)。

(2)育种。可用来固定杂种优势。对于一个杂合体 A/a 来说因为 A，a 会分离，所以不能作为种子，如果通过不等交换获得了顺式，那么就不会分离而可固定杂种优势，但这只能用于单基因控制的性状，而不能用于数量性状。

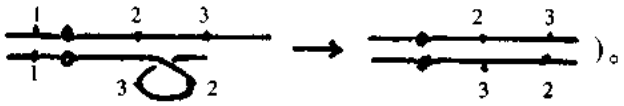
三、倒位

1. 倒位的类型

(1)**臂内倒位**(paracentric inversion)：倒位片段不含着称粒。()。

(2)**臂间倒位**(pericentric inversion)：倒位片段含着称粒。()。

2. 倒位的产生

(1)染色体纽结、断裂和重接()。

(2)转座因子可引起染色体倒位(见第 23 章)。

(3) 自发倒位，沙门氏细菌的相转变和 Mu 噬菌体的 G 片段的倒位分别是基因组中的 *hin* 和 *gin* 基因控制的，这种倒位属于正常的改变(见第 16 章)。

3. 遗传效应

(1)产生倒位环，抑制倒位区内的重组，形成假连锁(图 22－12)。无论是臂间倒位

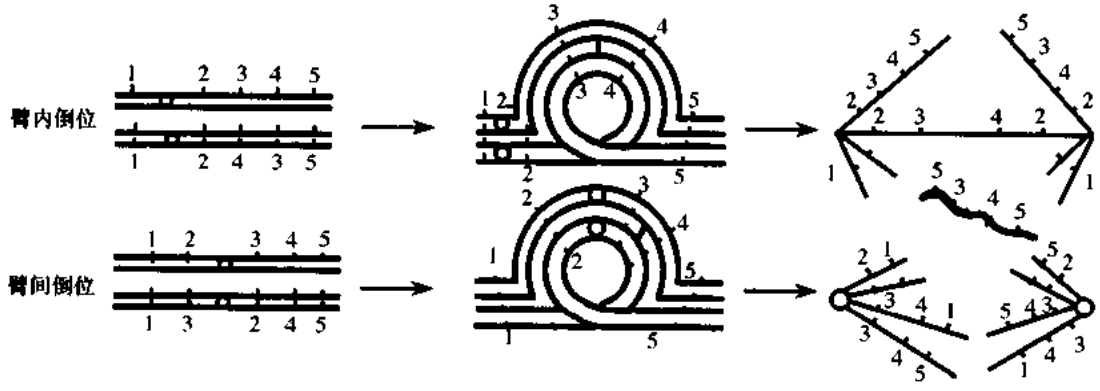


图 22－12　臂内倒位和臂间倒位的染色体在减数分裂时形成倒位环若倒位环内产生一次重组分裂时重组的染色体将产生缺失和重交，含有这种染色体的配子常常是无功能的。

还是臂内倒位在减数分裂时都会形成环(图 22 - 12,13)。若在环内发生一次重组,分裂时会产重复和缺失,含有这种重复和缺失的配子是没有功能的,这样在后代中就不会出现重组的类型,似乎是抑制了重组,实际上是重组的类型不能成活。

图 22 - 13　果蝇唾腺染色体臂内倒位环的照片。
(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 21 - 24)

(2)引起基因重排。倒位的结果引起了基因的重排,在原核生物中沙门氏细菌中由于 995bp 的片段倒位引起了不同鞭毛蛋白的产生,导致相转变。在 Mu 噬菌体中,G 片段的倒位产生了不同的 U.S. 蛋白,改变了宿主范围(见第 16 章)。

4. 倒位的应用

由于倒位可抑制重组,人们就利用此特点将它应用于突变检测和致死品系的建立。

(1)ClB 技术

早在 20 世纪 20 年代 Muller 就利用果蝇建立了一系列检测基因隐性突变和致死突变的技术,ClB 就是其中诸多方法之一,虽然现在已有一些先进的分子遗传学的手段但 ClB 技术仍不失为一种较为简便的方法,更值得借鉴的是 Muller 将基础理论应用于实践的一些思路,其构思精巧和灵活令人叹服。

ClB 技术中的 C 表示交换 (crossover),Muller 选择了 X 染色体上存在倒位的果蝇,目的是抑制交换,这样分析时就可以排除重组的因素。l 是指致死基因 (lethal,即在倒位的 X 染色体上已带有一个隐性致死基因;*B* 是指棒眼 (*Bar*) 基因,为显性突变。ClB 方法如下(图 22 - 14):将 ClB 品系的雌果蝇与待测的雄果蝇(例如事先经 X 光照射)交配,在 F_1 代中雌果蝇中应有两种表型:棒眼及非棒眼;在雄果蝇中应只有一种野生型可以存活,而另一种带有 ClB X 染色体的雄果蝇不能存活,因 Y 染色体上没有 *L* 基因的正常等位基因,使 *l* 表达而致死。再选择带有棒眼的 F_1 雌果蝇和野生型雄果蝇杂交,产生的 F_2 代也有两种类型,一种棒眼(ClB/+),另一种非棒眼(?/+)。棒眼雌果蝇一定带有一条 *ClB* 染色体,可以用来用下一轮检测。非棒眼的雌果蝇虽带有待测的 X 染色体,但由于另一条 X 染色体是野生型的,所有的基因都正常,所以即使待测的 X 染色体上发生了隐性突变或致死隐性突变都不显现出来。在雄果蝇中带有 ClB X 染色体的应不能存活,余下的一类雄果蝇是带有一条待测的 X 染色体和一条 Y 染色体,若经 X 射线照射产生了隐性突变即可在这一类雄果蝇中直接显现出来,若产生隐性致死,那么所有的雄果蝇都不能成活。

选择雄果雄作为待测的对象是因为:①一只雄果蝇可以和多只雌蝇交配,这样筛选突变(F_1 代)的样本就增加了。②雄果蝇是完全连锁的,不会发生交换。用 *ClB* 雌果

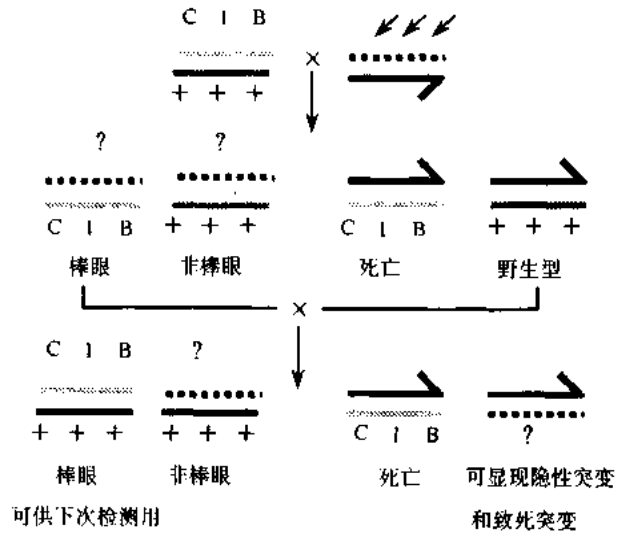


图 22－14　用果蝇进行检测隐性突变和致死突变的 ClB 技术。

蝇有以下的优点：①用 C(倒位）可抑制交换，使得结果明确，不受重组的干扰，②*B*（棒）眼作为一种标记易区分两类雌蝇，加将 *ClB* 品系分离出来；③*l* 使一半无待测 X 染色体的雄蝇致死，这样观察结果较为方便。

(2)保存带有致死基因的品系——平衡致死品系(balanced lethal system)

致死基因的存在可以供各种研究和检测(如 ClB 技术)用的，但获得一个致死突变并非易事，由于往往是隐性致死，即使是产生了也还得费很大的气力去检测和分离，因此人们总想建立一个致死品系，这样用起来就较为方便。首先提出平衡致死品系的设想是 Muller。1918 年他在实验室里发现了一种显性突变叫做蛎壳翅(beaded，Bd)(翅的前缘有扇形皱褶) Muller 将 *Bd* 突变的果蝇相互杂交，结果后代有两种类型，蛎壳翅和野生型，但比例不是 3∶1 而是 2∶1，表明 *Bd* 基因可能纯合致死，使 *Bd*/*Bd* 这种纯合类型致死。但有一次突然出现了意外的情况，蛎壳翅果蝇互交的后代都是蛎壳翅。Muller 对于偶然出现的现象并不放过，他对此做了一个假设，即在 Bd 等位基因的附近产生了一种新的致死突变，而且染色体存在着倒位，抑制了可能发生的重组。人们就利用此原理建立了一系列的平衡致死品系（图 22－15），这样它们互交的后代都是平衡致死的，即两个不同的致死基因反式排列在一对同源染色体上，且倒位的存在抑制了重组，无须选择就能保持真实的遗传，使致死品系得以保存。

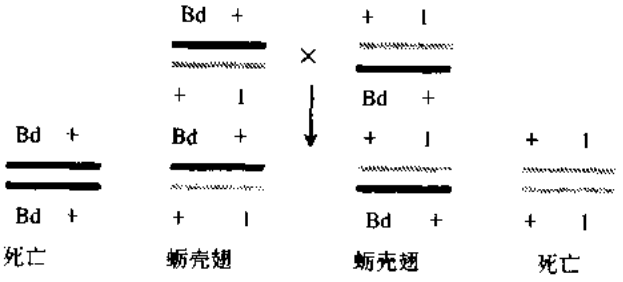
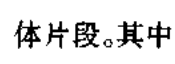
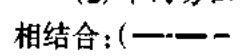


图 22－15　平衡致死品系的产生。Bd 为显性性状(蛎壳翅)，但仅纯合时方致死。

四、易位

1. 易位的类型

(1)相互易位(reciprocal translocation)　非同源染色体之间相互置换了一段染色体片段。其中又分为对称型()和非对称型(),非对称型染色体产生双着丝粒染色体及无着粒染色体片段,后者可形成微核或丢失。

(2)单向易位(simple translocation)　一条染色体的片段和另一条非同源染色体相结合:()。

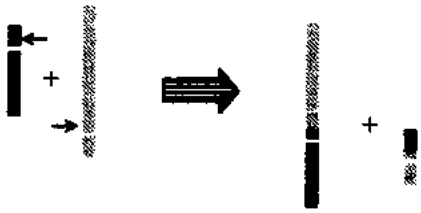


图 22-16　罗伯逊易位

(3)罗伯逊易位(Robertsonian translocation)。也称为**着丝粒融合或整臂融合**(whole-arm fusion)。这是一种交互非平衡易位。由两个非同源的端着丝粒染色体的长臂相互重接成一条长的中央着丝粒染色体,短臂相互连接成交互的交物,其上含有一些不重要的基因,这种小染色体常在细胞分裂中被丢失(图 22-16)。

(4)移位(shift translocation)　一条染色体片段移到同一条染色体的不同区域,或移到另一条染色体中称移位。前者称**染色体内移位**(intrachromosomal shift),后者称**染色体间易位**(interchromosomal shift)(图 22-17)。

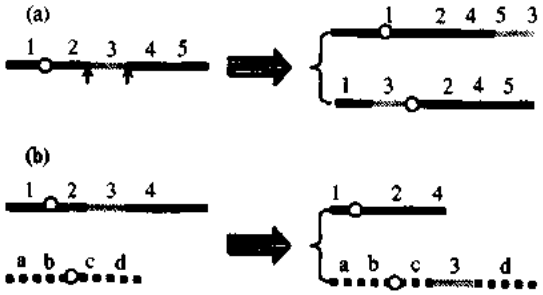


图 22-17　移位(a)染色体内移位;(b)染色体间移位。

2. 易位的产生

(1)断裂非重建性愈合。

(2)转座因子的作用(见第 23 章)。

3. 遗传效应

(1)位置效应

指由于位于常染色质的基因经过染色体重排移到异染色质附近区域,不能表达,表现出不稳定的表型效应。如玉米的胚乳是三倍体,若控制色素的基因是杂合的(C/c/

c)，当有的细胞中 C 基因移到异染区则不能产生色素，而大部分细胞却是正常的，使籽粒产生花斑。此和断裂－融合桥，体细胞突变及转座产生的花斑机制不同。同样当果蝇 X 染色体上的红眼基因 W^+ 易位到异染区也会出现红、白相间的复眼(图 22－18)。

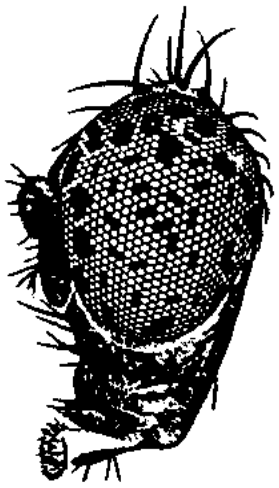


图 22－18 当果蝇 X 染色体上的红眼基因 w^+ 易位到异染区时也会出现红白相间的复眼。(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 8－34)

(2) 基因重排导致癌基因 (oncogene) 的活化产生肿瘤(见第 19 章)。

(3) 假连锁现象

当产生相互易位时，易位的非同源染色体在减数分裂中形成异常的“十”字配对 (图 22－19)。后期可能出现两种分离，一种是交互分离 (alternate segregation)，此分离是“十”字配对中不相邻的染色体同趋于一极，在减数分裂中期 I 形成“8”字形的四体环，同源染色体分向两极，产生了染色体正常的配子和平衡易位的配子。平衡易位虽有染色体的重排，但对整个配子而言基因组仍保持完整，没有重复和缺失。若无位置效应及重排则无严重影响，配子是可育的。这种分离的另一特点是末易位的染色体总是同趋向一极，易位的染色体上的标记基因如图中的 a，b 总是同时进入一个配子，而 A，B 总是进入另一个配子，形成了假连锁 (pseudolinkage)。另一类分离是邻近分离 (adjacent segregation)，即四体环中相邻的两条染色体同趋向于一极，这又有两种情况，一种是同源染色体同趋向一极，另一种情况是非同源染色体同趋向一极。不论哪一种情况其共同的特点是所产生的配子都存在重复和缺失，皆不可育，而自由组合产生的重组类型只存在邻近分离中，而它们的配子又都不可育。以上表明非同源染色体上的基因 a，b 或 A，B 各自都是不连锁的，由于邻近分离的产物不可育，只有交互分离的产物才可育，这样自由组合的类型都不会出现在后代中，而 a，b 或 A，B 原本不连锁的基因总是同时出现在一个配子中，似乎是连锁的，故称其为**假连锁**。

(4) 易位可导致人类家族性染色体异常

在人类中易位除产生肿瘤外还常引起各种家族性智力低下，其中也包括唐氏综合征。例如 t(7;9)(第七号染色体片段易位到第九号染色体上)。单向易位的个体减数分裂时可能产生四种不同的配子，和正常人结婚时可能产生四种后代 (图 22－20)：部分单体；部分三体；正常人和平衡易位的携带者。部分三体表型异常，智力低下。部分单体虽未在后代中发现，但可以预计对携带者的影响更为严重。因大片段的缺失往往配子难以成活或不育，故未能发现这种个体。平衡易位表型常常是正常的，但带有缺失和重复的染色体，和正常人婚配后，其后代同样出现以上四种类型，形成家族性染色体异常智力低下。平衡易位的个体即为携带者(图 22－20)。

如发生罗伯逊易位 (如 t(21; 14) 染色体罗伯逊易位) 那就可以有三种不同的分离方式，产生 6 种配子，和正常人结婚可产生 6 种不同的后代：21－三体即唐氏综合征，21－单体，14－三体，表型正常的携带者和正常人。携带者与正常人结婚又同样可产生 6 种不同的后代，形成了家族性的单体症、三体症(图 22－21)。

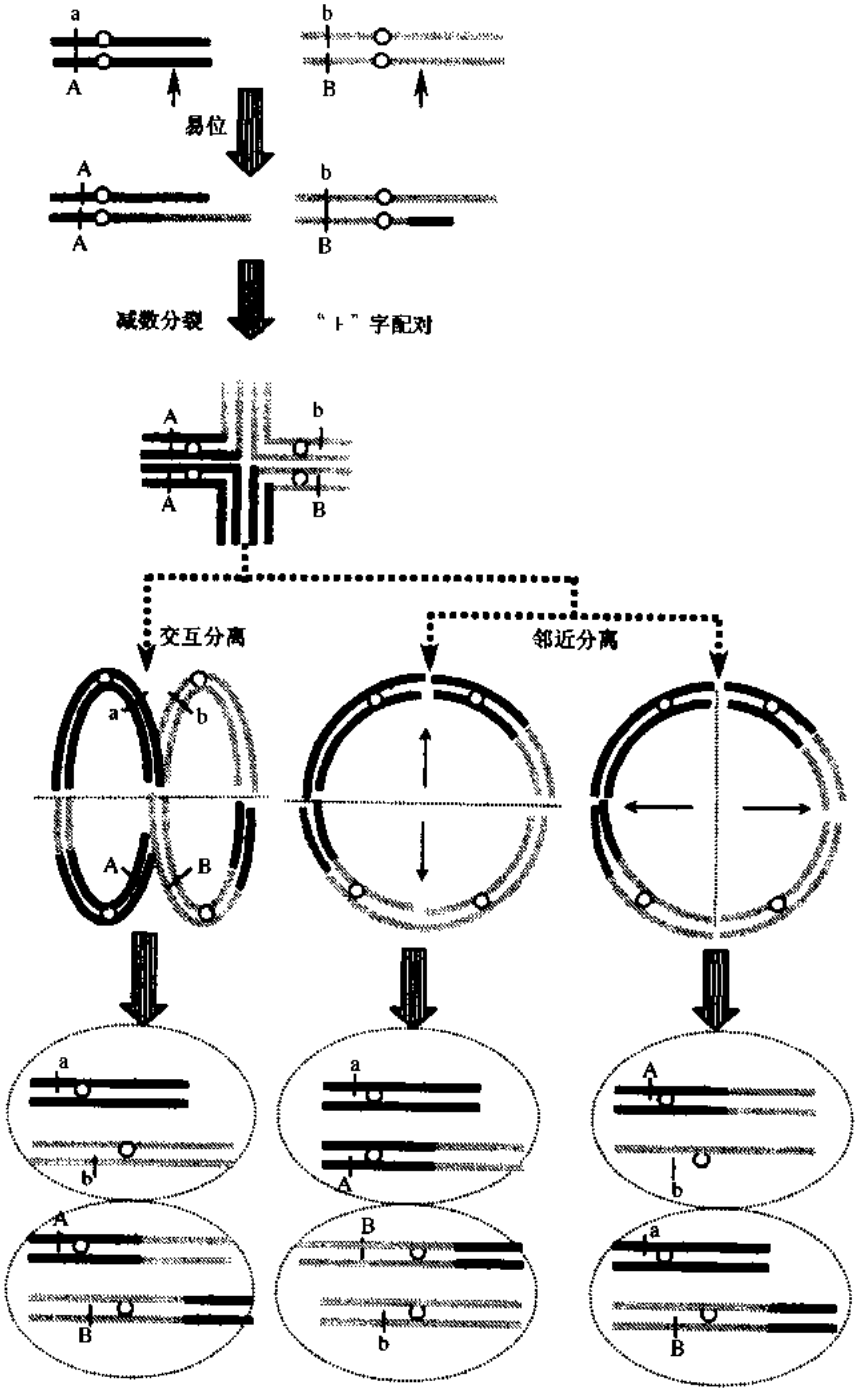


图 22－19　相互易位在减数分裂中形成了"十"字配对和四体环，不同分离的结果产生了假连锁。

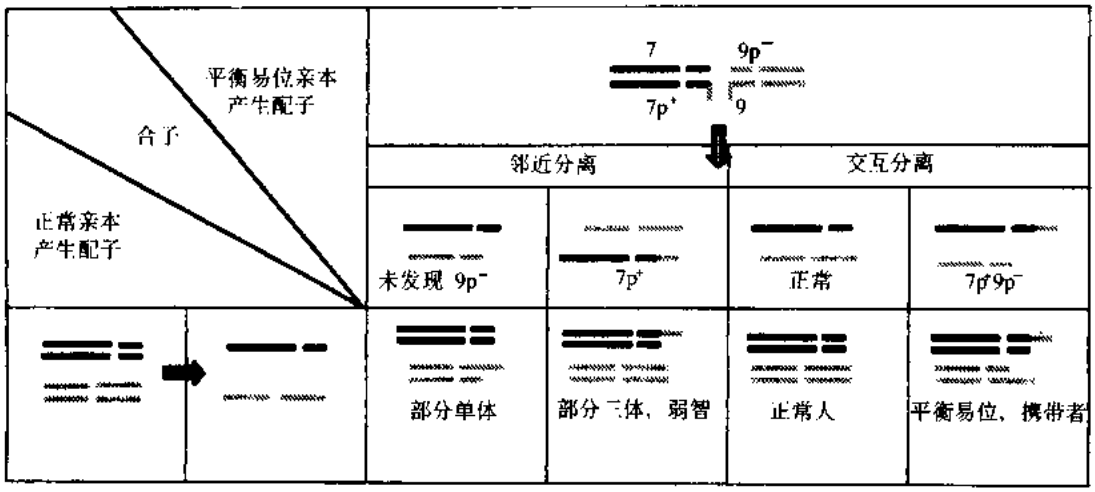


图 22－20　单向平衡易位的个体和正常人结婚，可能生育四种不同的后代。

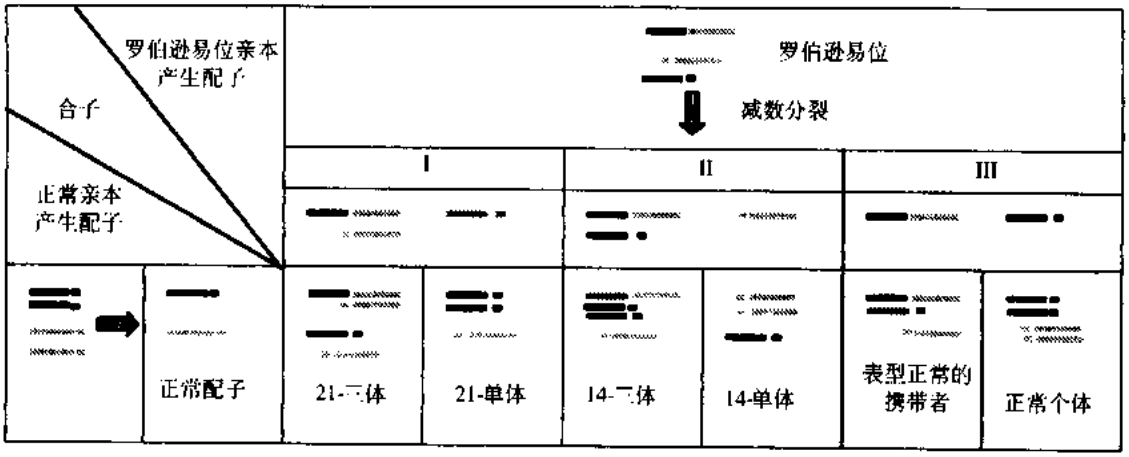


图 22－21　人类 21，14 染色体罗伯逊易位的个体其减数分裂有三种分离的方式，当和正常人婚配可能产生 21－三体，21－单体，14－三体，14－单体，携带者和正常人，形成了家族性染色体异常。

4. 易位、倒位与进化

在植物界中与进化较为密切是易位和倒位。它们是引起爆发式新种起源的重要途径。五十年代以来在钩瓣属(*Clarkia*)中发现许多种是以这种方式起源的(图 22－22)。*C. lingulata*(n = 9) 仅分布于加利福尼亚的默塞德河峡谷，此是另一种克拉花 *C. biloba*(n = 8)分布区的边缘；前者是自花受粉，后者是虫媒花。二者之间主要的区别是花瓣形状不同。*C. lingulata* 花瓣顶部无凹缺。两个种的杂交后代几乎完全不育，也不能回交。*C. lingulata* 不仅比 *Clarkia* 多出一对染色体，而且还存在易位和臂内倒位。Lewis 和 Roberts 认为 *C. lingulata* 是 *C. biloba* 在其分布边缘经过一系列染色体重建在近期内突然形成的。

开那克梯斯属的 *Chasenactis glabriuscula*(2n = 12)与生长在美国西南部较干旱的生态环境中的另外两个种 *C. fremontii*、*C. stevioides*(2n = 10)形态非常相似。但染色体数不同，种间杂种也是不育的。对种间杂种进行染色体组分析可以发现部分二倍体和三倍体，因此推测 *C. fremontii* 和 *C. stevioides* 都是由 *C. globriuscula* 通过染色体的易位和丢

失而派生出来的（图 22 - 22）由于染色体结构的差异使 3 个种产生了隔离而形成新种。在植物界中类似的例子还很多，在此不一一列举。

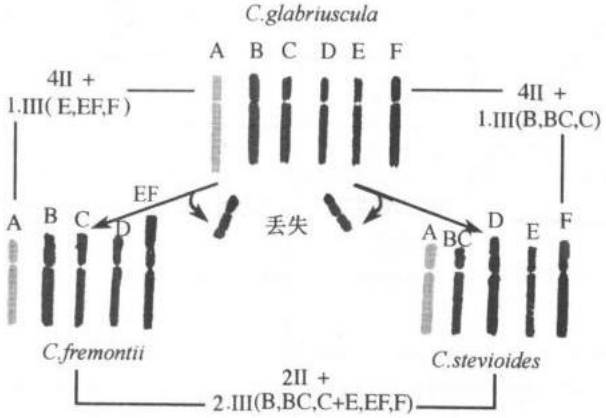


图 22 - 22　*Chasenactis glabriuscula* 和衍生种的染色体结构及种间杂种的减数分裂构型。（仿 Kyhos D. W: Evolution, 1965, 19: 26 - 43）

在植物界由于染色体分带技术尚不成熟，因此往往都是以观察杂种后代减数分裂中同源染色体的配对情况来加以分析、推测，手续比较繁琐，实验的周期也比较长，其结论也仅仅是假设，并不能得出十分明确的结果。而在动物界则不然，由于分带技术比较成熟因此无需进行杂交和染色体组分析，而只要将两个物种的带型进行比较就能得出明确的结论。例如人类（*Homo sapiens*）染色体 2n = 46；黑猩猩（*Pan troglodytes*）2n = 48；大猩猩（*Gorilla gorilla*）2n = 48；猩猩（*Pongo pygmacus*）2n = 48。人类的染色体核型与黑猩猩最相似，而和猩猩的差异最大。黑猩猩虽比人类多了 2 条染色体，但经 G 带比较发现人类 2 号染色体可能是黑猩猩丙条染色体着丝粒融合的产物，并且有臂间多次倒位

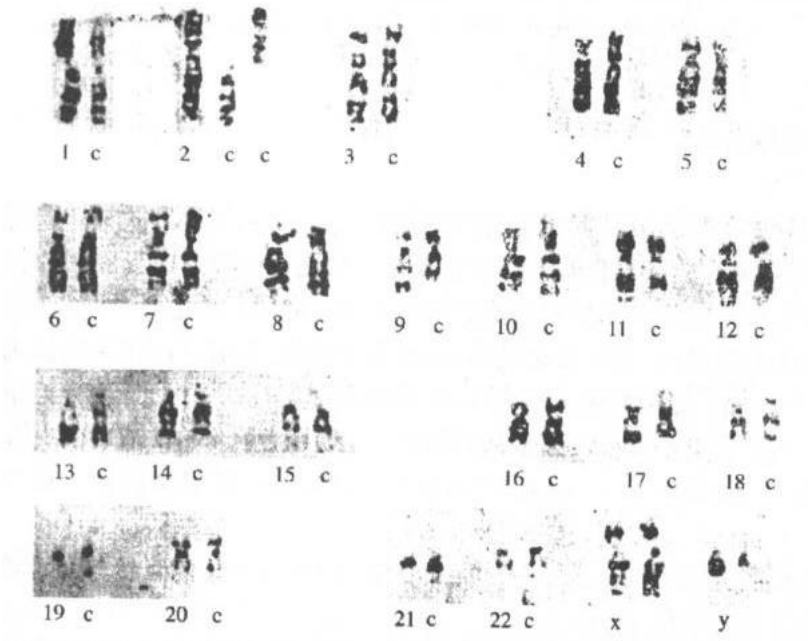


图 22 - 23　人和黑猩猩染色体比较。每一对的左边都是人的染色体。图中 C 代表黑猩猩的染色体。（引自 D. Warburton）

(图 22-23),其余的染色体都有很强的同源性,表明人类和黑猩猩的亲缘关系较近。

另一个有趣的例子是熊猫起源的争论,美国国家癌症研究所的 WilliamG.Nash 采用染色体分带技术对大熊猫和熊的染色体带型进行分析比较发现大熊猫的 1 号染色体可能是熊的 2、3 号染色体罗伯逊易位的产物;同样 2 号染色体与熊的 1、9 号染色体同源,3 号染色体与熊的 6、16 号染色体同源(图 10-28)。从而得出最后的分类建议,大熊猫与熊科中的熊有亲缘关系,小熊猫与浣熊科的浣熊有亲缘关系。

5. 易位的应用

(1)用于动物的育种

例如美国有一种优良葡萄品种,高产,糖分含量高,抗病力强,但果实比较小。为了改变这一性状保留其优点就采用放射线照其幼芽,使其产生易位,导致半不育。这样果实就会减少,相当于疏果的作用,使果实变大,而优良性状都保持如先。

(2)利用易位控制虫害

虫害的防治一直是件伤脑筋的事,因为用化学药品杀虫不仅价格贵,成本高,而且还会造成环境污染,另外防治的时间性也要求很强;利用天敌防治,问题在于天敌的数量难以控制,效果常不理想。

利用易位的方法可以有效地控制害虫。方法是以适当剂量的放射线照射雄虫,使其产生的后代约 50%不能孵化,出现半不育。应用此法要符合以下的条件:①这种昆虫的群体中是以一雄多雌的形式交尾;②在隔离区使用。中国农科院原子能研究所曾采用此法进行控制玉米螟的实验,方法是在玉米螟蛹孵化前两天用 2~2.5 万伦琴钴 60 照射蛹,羽化的玉米螟撒放到距内陆约 15 里的辽宁省菊花岛上,以期控制玉米螟。

据报道台湾柑橘果蝇也是采用此法,将幼虫进行放射线处理,蛹羽化后放入自然中,花了 10 年时间达到了控制的效果。

第三节 染色体畸变的鉴别

一、多线染色体的观察

双翅目昆虫摇蚊(*Chironomus*),果蝇和其他蝇类的幼虫唾腺里都具有多线染色体(polytene chromosome),它比普通染色体长了 100~200 倍,体积大了 1000~2000 倍,这是由于这些细胞十分特殊,在幼虫阶段 DNA 可复制 4~15 次,但细胞都不分裂,染色体丝平行排列,由于各段凝缩紧密程度不同,因而出现 5000 条以上明暗相间的带纹(图 22-24,25)。每斑带平均含有 30kb DNA。以前有很长的一段时间人们认为每个斑带只存在一个编码蛋白的基因,带间的区域存在基因间 DNA。实际上每条带编码多个蛋白。有时在一个带内的基因功能是相关的。

在多线染色体中同源染色体紧密配对,二者之间天衣无缝,因此一旦存在结构杂合体就会出现"环"的结构(图 22-6,13)。即使是结构畸变的纯合体通过斑带的对比也可以鉴别出来(图 22-11)。

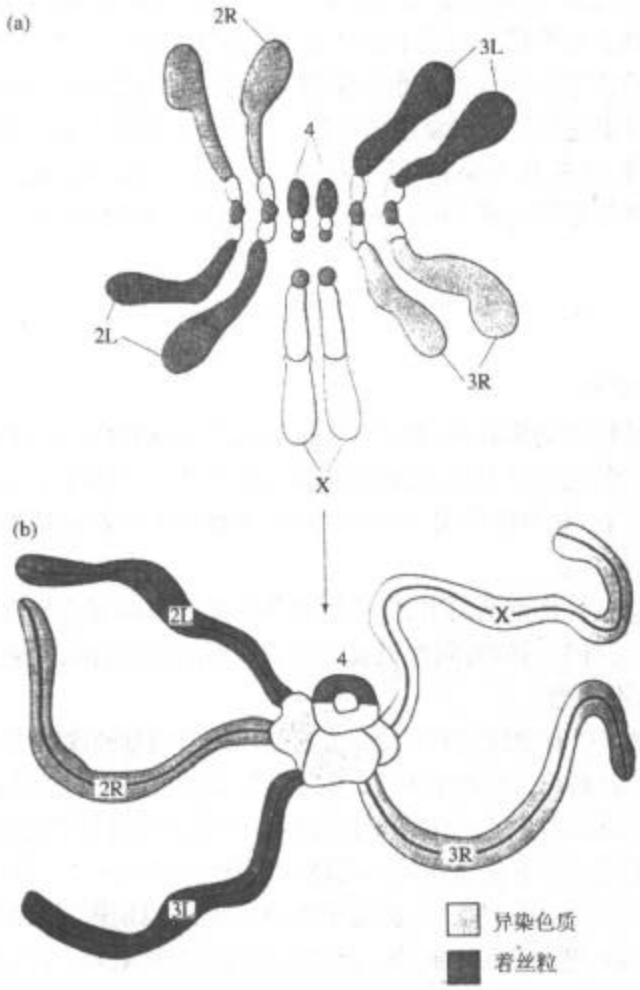


图 22－24　果蝇多线染色体的模式图。

（转引自 Griffiths A.J.F. *et al*.: An Intiroduction to Genetic Analysis, 5th ed., 1993, Fig. 8－3a, b）

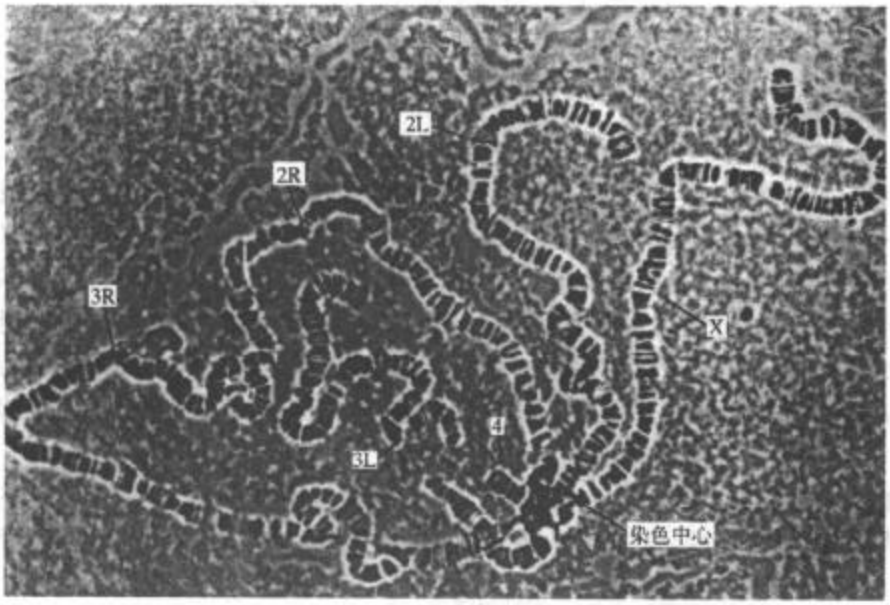


图 22－25　果蝇多线染色体的显微照片。

（转引自 Griffiths A.J.F. *et al*.: An Intiroduction to Genetic Analysis, 5th ed., 1993, Fig. 8－3）

二、染色体带型的比较

动、植物的核型可以鉴别出明显的染色体畸变，如重复，缺失，易位，双着丝粒染色体，环状染色体，额外小染色体等，但不能确定具体的易位片段来自哪一条染色体。由于制片过程中染色体的大小，长短会产生变化。因此稍长一点或稍短一点并不能确定是否是重复、缺失或易位，对此须持谨慎的态度。而采用分带技术就可以明确地判断各种畸变，包括倒位（图 22－4，22－19，11－28），但涉及到很小的片段仍不能用这种方法。

三、用微铺展法制备联会复合体

微铺展（microspreading）法是 20 世纪 80 年代建立的一种技术，这是用减数分裂的精母细胞悬液滴加在涂有 Falcon 塑料膜的载玻片上，涂匀干燥后固定，经硝酸银染色，将带有分裂相的 Falcon 膜切下浮在清水上，移置光镜下观察，寻找联会复合体（synaptonemal complex，SC）清晰区域，用单孔铜网套住，放在水上，移置电镜下观察、拍照（图 22－26），可见到铺开的 SC。若存在缺失、重复、易位和倒位就能清晰地观察到单股环（缺失或重复产生的不配对区）和双股环（倒位环）（图 22－27）。

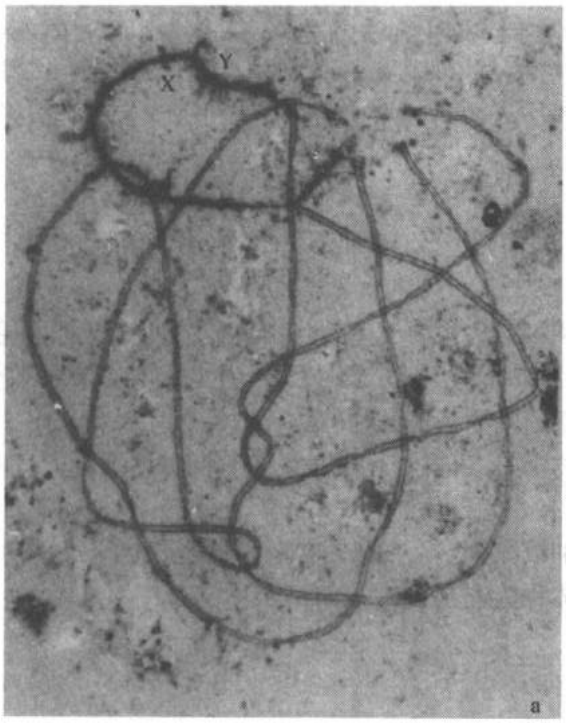


图 22－26　用微铺展法制备的赤麂（*Muntiacus muntjac*）的联会复合体。（照片由史庆华博士惠赠）

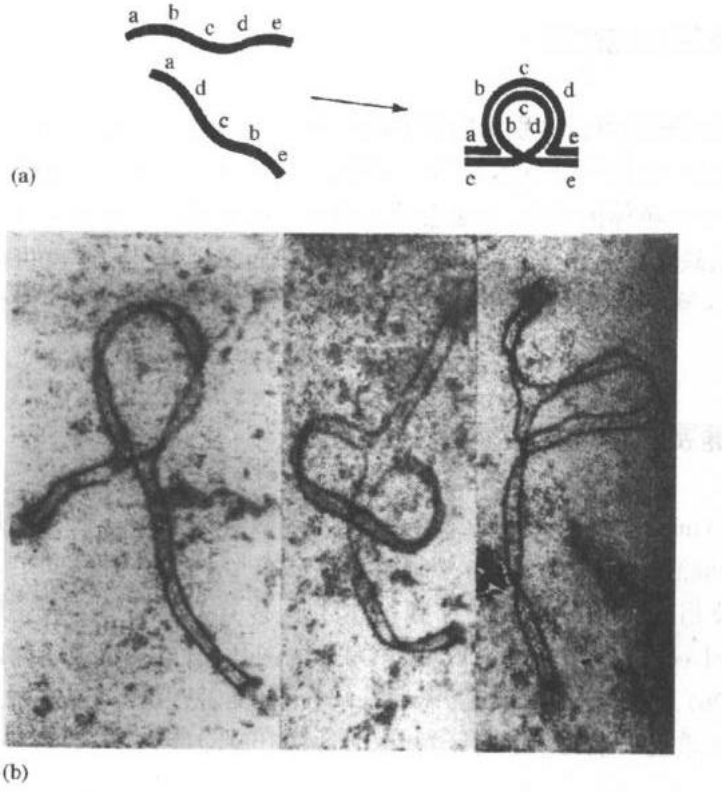


图 22 - 27　杂合配对的染色体在减数分裂中形成倒位环。(a)倒位环的模式图,图中的各条染色体实际上是姊妹染色单体。(b)用微铺展法制备的杂合小鼠的联会复合体电镜照片。

(转引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 8 - 18)

第四节　染色体数目的改变

染色体突变的和第二种主要类型就是染色体数目的改变。染色体数目的改变对人类的影响也是很大的。在植物中也会自发地产生染色体数目的改变。其实染色体数目的改变在进化中对塑造基因组是有利的。

大部分真核生物都是二倍体,也就是说在它们的体细胞中有两套染色体。而它们的配子是单倍体,只有一套染色体。

二倍体生物增加或减少整套的染色体就称为**整倍体**(enpeoid);若增加或减少一条或几条完套的染色体称为**非整倍体**(aneuploid)

一、整倍体

一套基本的染色体的数目称为**一倍体数**(x)。带有多个一倍体染色体数(monoploid number(x))的生物称为**整倍体**。整倍体的类型分为:**多倍体**(polyploid):具有两套以上的染色体。这样 1x 是一倍体,2x 是 2 倍体(diploid)。多倍体指三倍体(triploid)(3x)、四倍体(tetraploid)(4x)、五倍体(pentaploid)(5x)、六倍体(hexapliod)等。单倍体数(n)是指配子中的染色体数。大部分动、植物的单倍体数和一倍体数是相同的。n 或 x 可以交替使用。但在某些植物如小麦其 n 和 x 的意义就不同。小麦有 42 条染色体,共有 6 套,因

此它是 6 倍体，也就是说 2n = 42 = 6x。但配子里有 21 条染色体，所以单倍体含 n = 21 = 3x，2n = 42。

1. 一倍体

雄性蜜蜂，黄蜂和蚁都是一倍体（见第 24 章）。在这些昆虫的生活周史中雄性是由孤雌生殖发育而成，即由未受精的卵发育而成。在大部分物种中一倍体个体是不正常的，在自然群体中很少产生这种异常的个体。

一倍体的精细胞不是经过正常减数分裂产生，因为它们的染色体只有一套，不存在同源染色体的配对。这样雄性是不育的。（雄蜜蜂、黄蜂和蚁在形成配子时绕过减数分裂，通过有丝分裂产生配子）。如果一倍体细胞进行减数分裂的活，单独一条染色体随机趋向两极，那么所有的染色体都趋向一致的概率是 $(1/2)^{x-1}$，x 是染色体数，这将决定有功能配子（含全套染色体）的频度，很显然，如 x 较大的活可育配子产生是极少的，几乎为零。

在现代植物育种中常应用单倍体培养技术，如花粉培养技术。花粉是单倍体，通过冷处理的诱导能培养成胚状体（一种小的可分裂的细胞团），经进一步培养可长成单倍体植物。

一倍体植物对于筛选某些隐性的抗性突变十分方便，只要将一倍体细胞放在选择性培养基上就可筛选出抗性细胞，然后培养成抗性一倍体植株，但无繁殖能力，用秋水仙碱适当处理，使染色体加倍，便可获得纯合的抗性可育的二倍体植株。80 年代我国在花粉培养方面做了很多工作，在国际上处于领先水平。

2. 多倍体

我们涉及到多倍体这一领域就必须区分**同源多倍体**（autopolyploids）和**异源多倍体**（allopolyploids），同源多倍体的多套染色体来源于同一物种，而异源多倍体的多套染色体来源于不同物种，但这些物种的亲缘关系很近，它们整套染色体之间是部分同源的（homeologous），而不像同源多倍体是完全同源的。

3. 三倍体

三倍体通常是同源的，是由 4x（四倍体）和 2x（二倍体）之间自然或人工杂交而产生的，2x 配子和 1x 配子结合形成三倍体。

三倍体的特点是不育，这和减数分裂时染色体分离有关。无论是同源三倍体还是异源三倍体在减数分裂的后期每一号染色体中都有一条染色体可随机的趋向于一极，只有各号数染色体中的这一条染色体同趋向某一极才能获得染色体平衡的配子，否则是不育的。因此产生有效配子的概率为 $(1/2)^{(x-1)}$。这个可能性很小，故无论是同源三倍体还是异源三倍体都是不育的。

香蕉就是同源三倍体水果，每套染色体含 11 条染色体，产生有效配子的概率是 $(1/2)^{(11-1)} = 1/2048$，也就是说约在五百万次机会中仅有一次可以由两个有效配子形成合子，因此香蕉是不育的，香蕉中没有种子。

中国水仙（*Narcissus tazetta L. var. chinensis Roem*）是我国很普及的一种观赏植物，

它也是三倍体植物，x = 10, 3x = 30(图 22 - 28)，水仙是三倍体，故只开花不结实，一般依靠鳞茎营养繁殖。

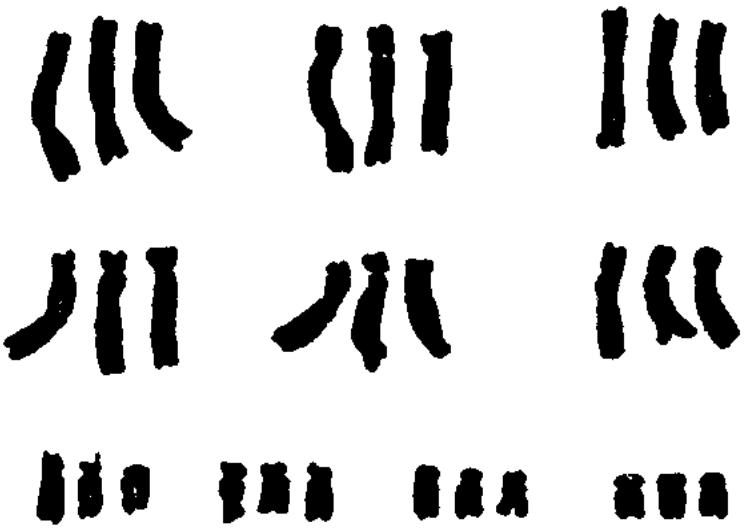


图 22 - 28　中国水仙的核型(3x = 30)。(引自朱心武等：武汉植物学研究，1986，(4)2)

4. 同源四倍体

同源四倍体是自然产生的，一般是 2x 基因组通过染色体自发加倍而产生 4x 的基因组。也有的是通过用秋水仙素人工加倍而产生。同源四倍体植物作为一种商品化的作物具有一定的优势。因为多套染色体常使植株高大、细胞、花、果实都比二倍体的要大一些。

由于 4 是偶数，所以同源四倍体可能有规则的减数分裂，但并非总是如此，关键在于要看每套染色体(4 条)如何配对和分离。两条同源染色体相互配对的叫**二价体**；3 条同源染色体配对叫**三价体**（trivalents）；4 条同源染色体配对的叫**四价体**（quadrivalents），而一条不配对的染色体叫**单价体**(univalent)。在四倍体中，二个二价体或一个四价体的配对形成可以产生正规的分离(图 22 - 29)，但并不保证 2 - 2 分离。在某些物种

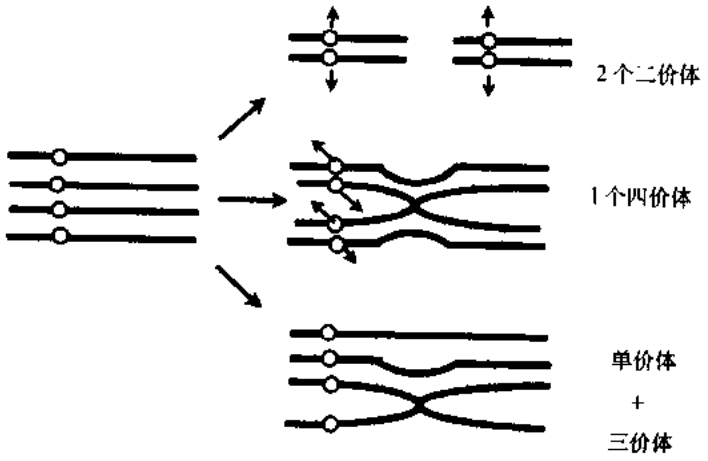


图 22 - 29　同源多倍体减数分裂时 4 条同源染色体可能的配对形式及分离。

它们是呈 2－2 分离，那么所有的配子都是有功能的，这样的 4 倍体是可以正常发育的，如四倍体曼陀罗(*Datura stramonium*)(4x = 48)和番茄(4x = 48)。

现在让我们来考虑一下可育四倍体的遗传学。如果我们用秋水仙素使基因型 Aa 植物的染色体加倍形成 AAaa 四倍体，假设这四倍体是 2－2 分离，那就会有几种不同的配对情况，在它们的后代中可能产生不同的分离比，取决于此基因的座位是否和着丝粒紧密连锁。首先我们考虑一个着丝粒基因，它有 3 种可能的配对和分离方式，在配子中染色体有不同的组合(图 22－30)，但发生各种情况的几率相同，因此所产生 3 种基因型(Aa，AA 和 aa)的配子，其比率是 8∶2∶2 或 4∶1∶1。这样的植物如果自交的话产生 aaaa 后代的概率是 1/6×1/6＝1/36。若 A 对 a 是完全显性的话，F_2 代的分离比不再是 3∶1，而是 35∶1。

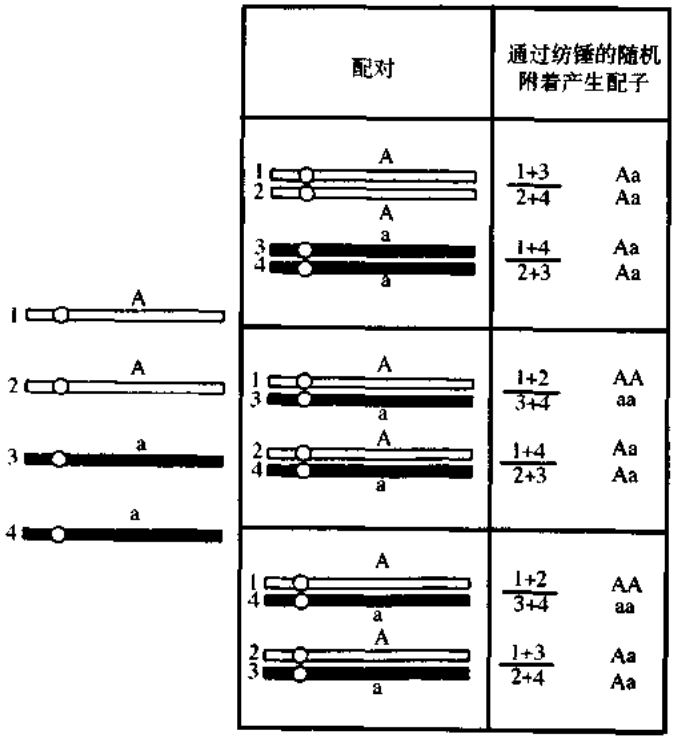


图 22－30　在四倍体中基因的分离，假设这四倍体是 2－2 分离，那就会有几种不同的配对情况。
(转引自 Griffiths A. J. F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 9－7)

如果同一种植物中等位基因 B 和 b 的座位离着丝粒很远，那么必须要考虑到交换。为了讨论交换我们就要涉及染色单体。假设有 4 条 B 染色体单体和 4 条 b 染色体单体，在 B/b 与着丝粒之间的长区域中发生多次交换，那么两个基因进入同一个配子的情形象是在装了 8 粒围棋子(4 黑 4 白)的袋子中我们随机地每次摸出 2 粒一样，结果会是 4 种相同(B/B)，或(b/b)，4 种不同(B/b)。那么两个 b 基因先后进入一个配子的机会是

4/8(第一次 b 进入的概率)×3/7(第二次 b 进入的概率)＝12/56＝3/14；自交产生 bbbb 表型的概率是 3/14×3/14＝9/196≈1/22。因此表型 BBBB∶bbbb＝21/1。

5. 异源多倍体

经典的异源多倍体是由 G. Karpechenko 于 1928 年合成的。他想培育一种可育的杂种蔬菜，地上部分长有甘蓝（卷心菜 *Brassica*）的叶子而地下部分长萝卜（*Raphanus*），甘蓝和萝卜这两个物种都是 2n = 18，而且亲缘关系比较近，可以进行杂交。从种子获得可以生存的杂种后代。但由于两亲本染色体数虽相同但不同源，所以不能正常地配对联合，因此杂种是不育的。但有一天这种不育的杂种却产生了几粒种子，这些种子播种后长成了带有 36 条染色体的可育个体。 这些个体都是异源多倍体。它们是由于不育杂种染色体自发的加倍而产生了 $2n_1 + 2n_2$。这种个体的染色体可配对和正常分离，产生有功能的配子，这种异源四倍体也称为双二倍体（amphidiploid）。萝卜甘蓝虽然和预期的目的正相反长出了萝卜的叶子和甘蓝的根，但仍很有意义。因为这是人为创造出的一个新种叫萝卜甘蓝（*Raphanobrassice*）。将它和两个亲本分别回交所产生的后代都是不育的。

现代，异源多倍体已是植物常规育种的一种手段，人们用秋水仙素加倍染色体取代自发加倍。育种者的目的是将两个亲本的优良性状进行重组，这是用传统杂交的方法所不能达到的。例如小黑麦（*Triticale*）双二倍体是由普通小麦（*Triticum*， 2n = 6x = 42）和黑麦（*Secale* 2n = 2x = 14）重组而成。小黑麦具有小麦高产和黑麦的耐瘠、耐寒的特点。一个巨大的国际小黑麦测试项目已在进行，很多育种学家对于这种人工双二倍体的特点寄予厚望（图 22 - 31）。我国学者鲍文奎利用含有可杂交基因的桥梁品种先与

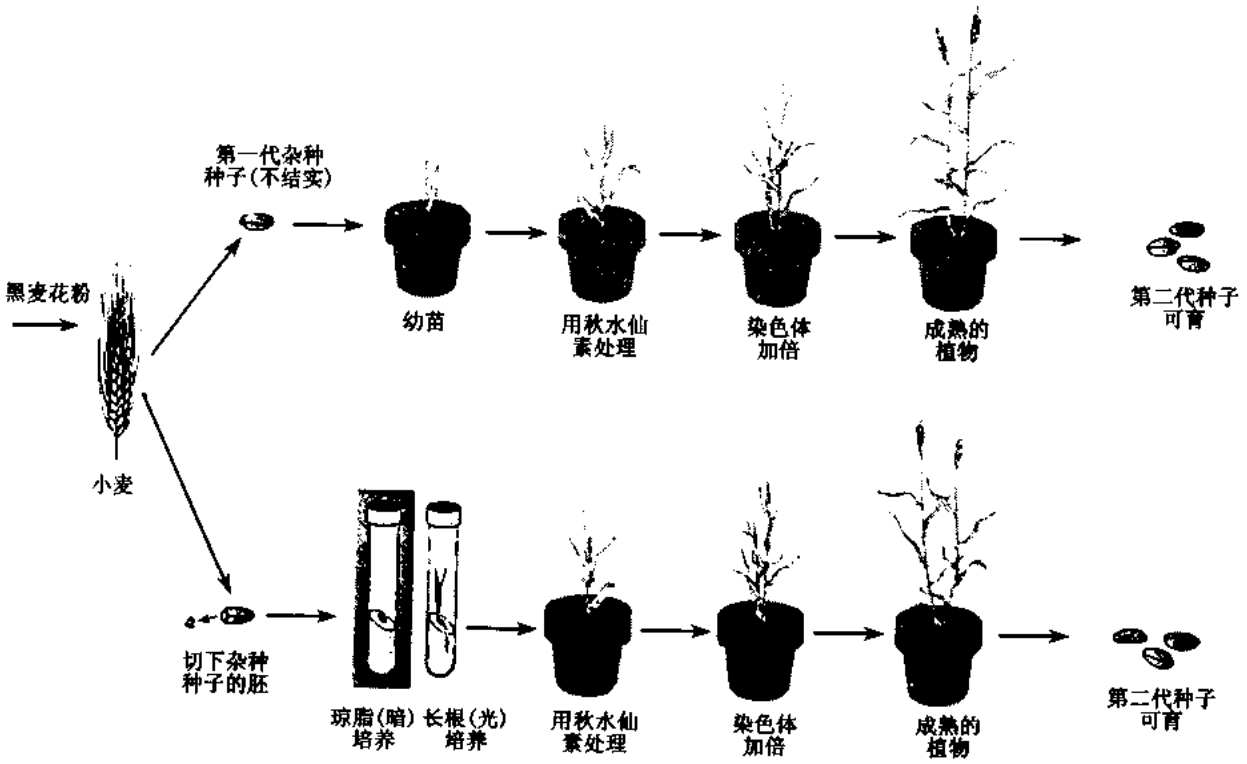


图 22 - 31　双二倍体小麦育种技术。若这种杂种种子不能萌发，那么通过组织培养可以获得杂种植物。
（转引自 Griffiths A. J. F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 9 - 10）

小麦杂交，让杂交后代获得可杂交基因，再与黑麦杂交，这样解决了小麦与黑表属间杂交的困难，成功地培育了 8 倍体小黑麦。

在自然界中，异源多倍体看来主要发生在植物中，已有很多不同的例子，如芸苔属（*Brassica*）植物中就有异源多倍体的存在（图 22－32）。它们有 3 个不同的基本种，通过杂交重组产生了新的双二倍体。

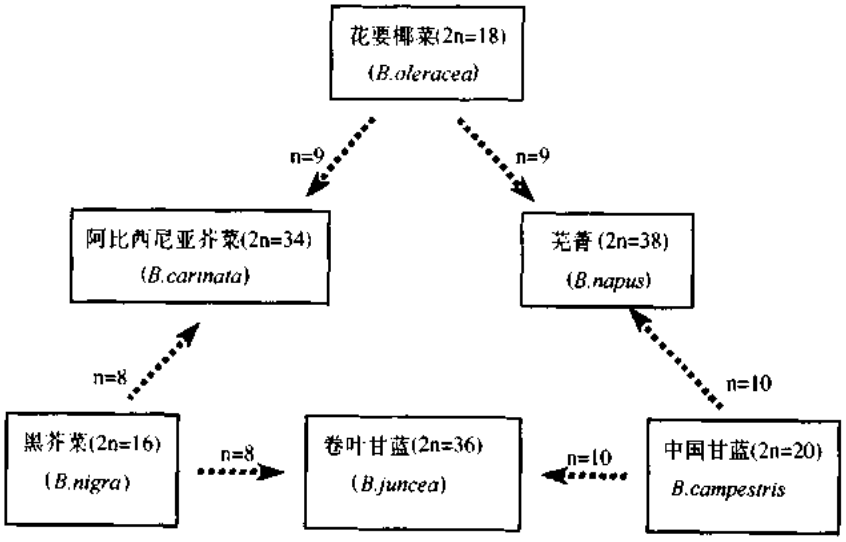


图 22－32　芸苔属植物中由三个亲本种相互杂交，经染色体加倍形成了不同的四倍体新种。
（仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 9－1）

人们通过染色组分析（图 22－33）（即通过对杂种后代减数分裂中可能形成的二倍体和单倍体数目来判断两个亲本染色体的同源性，以推测它们之间的亲缘关系），对普通小麦的起源作了如下的推测：两种野生二倍体小麦的杂种（AB）偶然经过自发地染色体加倍，形成了野生的四倍体小麦（AABB），经长期栽培驯化形成了二粒小麦。二粒小麦和另一种野生的二倍体小麦杂交，杂种后代经偶然加倍形成了现代的 6 倍体普通小麦。

多倍体的植物是常见的，但多倍体动物就十分罕见了，但也发现一些例子，如扁

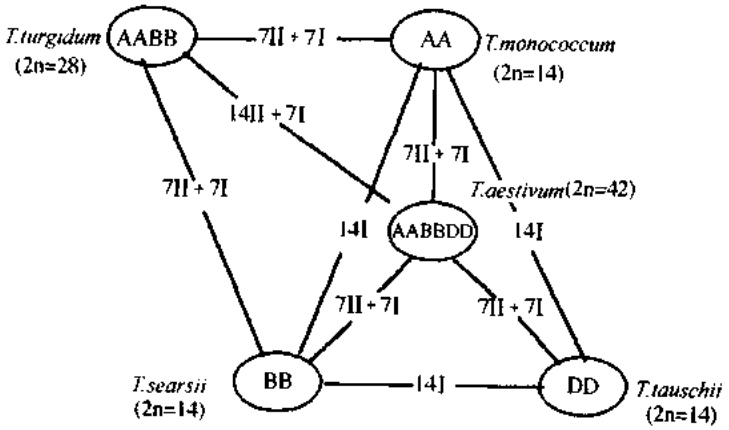


图 22－33　小麦属（*Tyiticum*）中，几种野生小麦和普通小麦的染色体组分析。

形虫，水蛭和海虾中都有发现。它们是通过孤雌生殖的方式繁殖。特殊类型的未受精卵发育成胚而无需受精。多倍体的例子不只是出现在低等动物中。在鱼类，两栖和爬

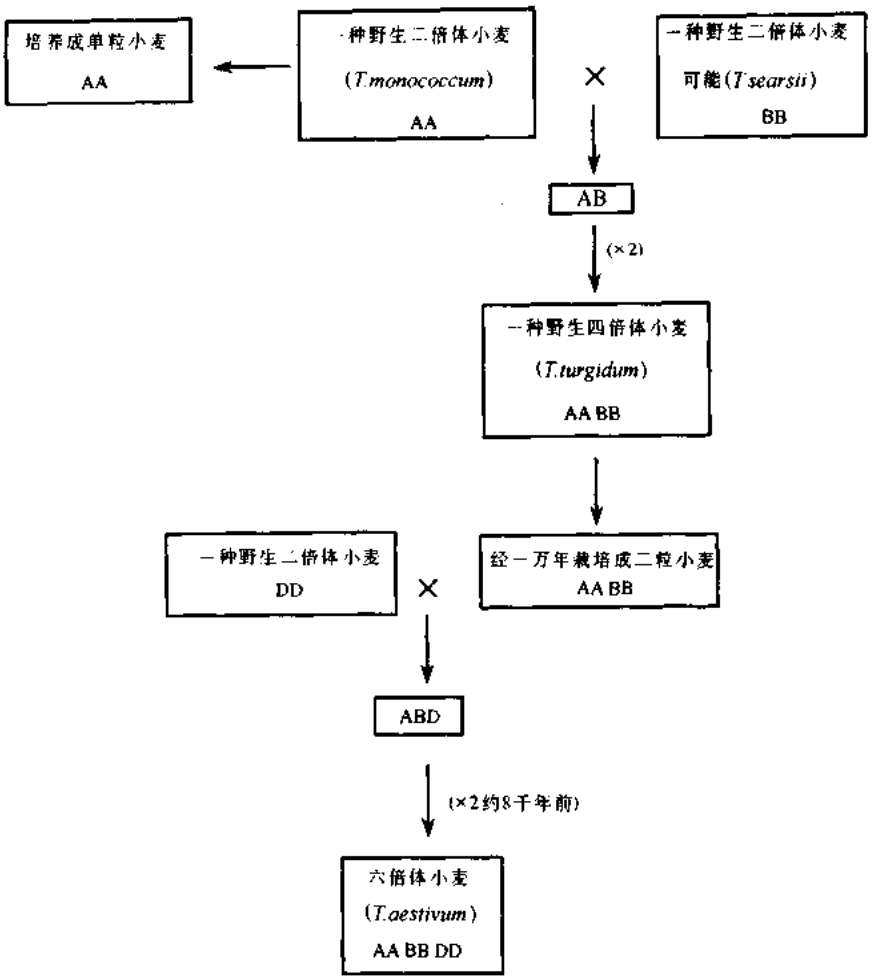


图 22－34　现在的六倍体小麦可能的演化途径。

(引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 9－12)

行动物中也都存在多倍体。它们有各种繁殖的方式。多倍体蛙和蟾蜍雌雄个体可进行有性生殖,而多倍体蝾螈和蜥蜴进行孤雌生殖。

某些鱼类也有多倍体。有的是由单个的多倍体在进化中产生了个完整的分离群。鲑科的鱼类(包括鲑属和鳟属)是一开始就多倍化的例子。鲑属鱼的 DNA 含量是相关鱼类的 2 倍，不同种的鲑鱼具有不同数目的染色体，但染体臂数不变(在有些种内有染色体融合)，在相关的种群中有 2 倍的臂数。证据表明鲑鱼是由单个的多倍化发展而形成四倍体。

有趣的是在动物中也存在不育的三倍体。三倍体牡蛎是存在的,而且比相关的二倍体更具有商业价值。二倍体进入产卵季节时味道不好,而三倍体是不育的,不产卵,因此一年四季味道都很鲜美。

人类的多倍体受精卵可因细胞分裂发生差错而产生，但大部分在胚胎期死亡，偶然会降生一个三倍体的婴儿,但无一例成活。

二、非整倍体

少数的二倍体生物增加或减少了一条或几条完整的染色体，而不发生整套染色体的增减，称为**非整倍体**。在大部分情况下在动物中非整倍体是致死的。非整倍体在植物中发生的频度是较高的，可能由于在植物中非整倍体常得以生存之故。其产生的原因可能是在有丝分裂或减数分裂中个别染色体的不分离或丢失所致。

正常的个体是完整的二倍体（2n），又可称为**双体**（disomy）在此基础上我们可将非整倍体分成以下几种类型。

（1）**缺体**（nullisomy），2n－2，个体丢失了一对同源染色体，又称为**零体**，一般是致死的，因丢失染色体上带有的基因是其他染色体所不具有的，无法补偿其功能。但在异源多倍体植物中常可成活，但长得较弱小。

（2）**单体**（monosomy），2n－1，个体丢失某一条染色体，虽然丢失染色体的同源染色体存在，但由于以下的原因单体仍出现异常表型或特征。①染色体的平衡受到破坏；②某些基因产物的剂量减半，有的会影响性状的发育；③随着一条染色体的丢失，其携带的显性基因随之丢失，其隐性等位基因得以表达。

（3）**三体**（trisomy），2n＋1，个体多了某一条染色体，由于染色体平衡的破坏和基因产物剂量的增加，三体也显示出异常表型特征。在人类中常见有三种三体：①21－三体即 Down 氏综合征；②18－三体，即 Edward 综合征；③13－三体，即 Patau 综合征。

（4）**双三体**（double trisomy），2n＋1＋1，即个体多了两条不同的染色体。

（5）**多体**（polysomy）2n＋n，某一号染色体增加一条以上称为多体。

21－三体病又称为**先天愚型或唐氏综合征**（Down syndrome）它是由英国医生 Langdon Down 首先描述其症状，故称为 Downs syndrome。1959 年法国遗传学家 Lejeune 发现此病是由于多了一条 G 组染色体所致，后确定这条染色体是 21 号染色体，故此病又称为 21－三体综合征（图 22－35）。Lejeune 的发现开创了医学遗传学的一个分枝——临床细胞遗传学。

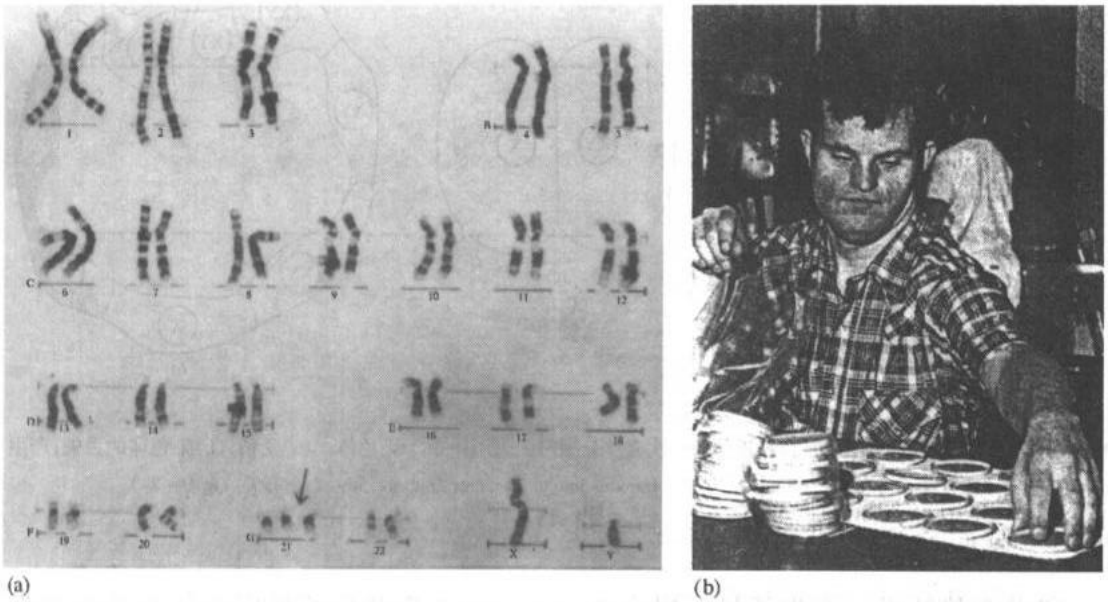


图 22－35　21－三体（唐氏综合征）（a）核型；（b）患者。（引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 19－14）

患者鼻扁平，眼裂向外上倾斜，眼距过宽，嘴小唇厚，舌大耳小，手宽而肥大，常为通贯手，智力低下。新生儿发病率为 1.5/‰。其产生的原因可能是母亲卵细胞形成过程中染色体的不分离，或发生罗伯逊易位所致。

由于女性个体的减数分裂是从本身处在胚胎期就开始了，到性发育成熟要经历漫长的时间，这样染色体发生畸变的机会也较大。母亲生育年龄的增加，21－三体的发生率也随之增加(表 22－2)。因此晚婚晚育应符合自然规律，要有一个适当的度，决不能以降低人口的素质为代价来换取出生率的降低。

表 22－2　母亲年龄和 21－三体发生率的关系

母亲的年龄	21－三体发生率
< 29	1/3000
30－34	1/600
35－39	1/280
40－44	1/70
45－49	1/40

摘自 *Genetics*, 3d. ed 1992. Peter J. Russell

三、体细胞非整倍体

在体细胞或培养细胞中也会产生非整倍体。例如嵌合体(genetic mosaic)就是在发育早期产生了非整倍体。人类中的某些异常也提供了很好的例子。

两性嵌合体人既具有男性内生形殖腺睾丸，又具有女性的卵巢。这是人类第一例体细胞非整倍体。有一种类型的两性嵌合体为 xo/xyy，此可能是在 xy 合子发育早期，在有丝分裂中两条 y 染色体的姐妹染色体单体没有分离，同趋于一极，使另一极缺少了 y 染色体。这样，一个子细胞的后代为 xyy，另一个子细胞的后代为 xo 。这种个体的表型性别取决于身体的某一组织是由什么样类型细胞克隆组成的，是 xyy 型，还是 xo 型。如果不是在受精卵一开始分裂就产生染色体不分离的话，就可能产生三种类型的嵌合体 xy/xo/xyy。它含有一个正常男性细胞的克隆。另一种两性嵌合体为 xo/xy。此可能是 xy 合子在发育早期的有丝分裂中丢失了一条 y 染色体所致(图 22－36)。真两性畸形中的 xx/xy 可能是两个受精

256

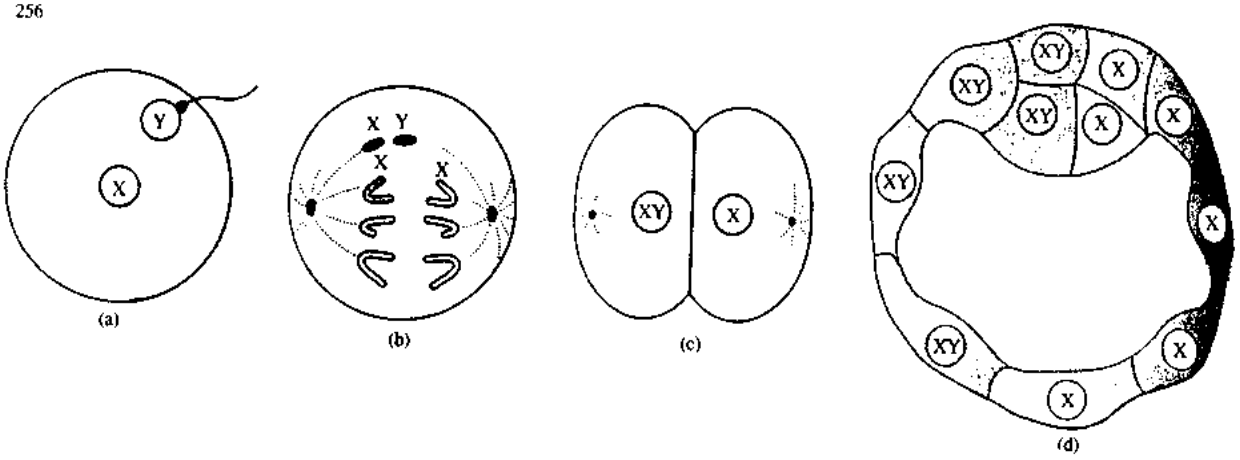


图 22－36　人类两性嵌合体 XY/XO 是由于在合子进行第一次有丝分裂时丢失了一条 Y 染色体。(a)受精；(b)染色体的丢失；产生了两性嵌合体。(引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 9－24)

卵融合的结果，产生了异源嵌合体 (chimera)。通常称两性嵌合体为**两性畸形**(gynandromorphs)。在动物中也发现两性畸形的存在。一个典型的例子就是果蝇。在果

蝇的 x 染色体上连锁着控制红眼的基因(w^+)和长翅基因(m^+)。它们的等位基因分别是 w(白眼)和 m(小翅)。一个 w^+m^+/wm 杂合体,如果在第一次有丝分裂丢失了含有 w^+m^+ 的 X 染色体,那么由这个细胞产生的克隆的基因型是 o/um,表型是白眼小翅,而且因只有一条 X 染色体,性指数为 0.5,故其性别表型应为雄性,而未丢失 X 染色体的子细胞将发育为正常的细胞克隆,基因型为 w^+m^+/wm,表型为红眼,长翅。因有两条 x 染色体,性别表型应为雌性。结果产生了两性嵌合体。从身体的中线为界,一半发育为雄性,白色眼,小翅;尾部有三条黑色斑带;另一半发育为雌性,红色眼长翅,尾部有五条细的斑带(图 22 - 37)。

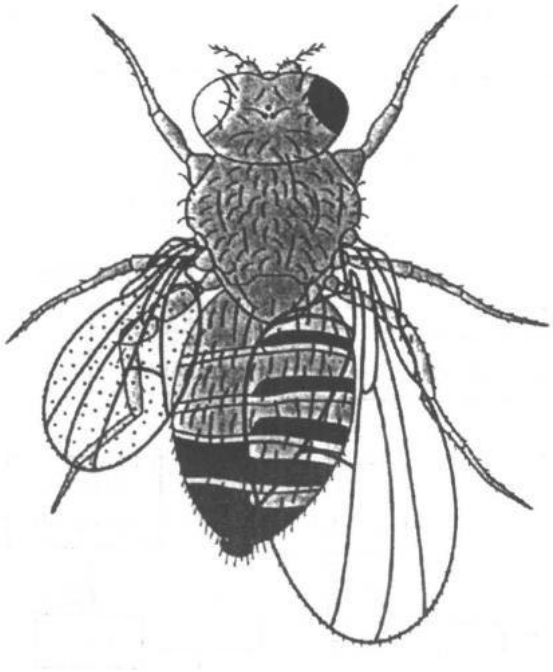


图 22 - 37　果蝇的雌雄嵌合体。此合子是 w^+m^+/wm,当第一次卵裂时一个细胞丢失了带有 w^+m^+ 基因的 X 染色体,发育成两性嵌合体。身体的一半为雄性(左),白眼,其基因型为 $O/\ wm$;身体的另一半为雌性(右)红眼,其基因型为 w^+m^+。(引自 Griffiths, A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 9 - 25)

体细胞非整倍体和其产生的嵌合体往往和癌也有关。前面已介绍了费城染色体是 22 号染色体长臂易位到了 9 号染色体上。但在慢性髓性白血病 (chronic myeloid lenkemia, CML)患者中发现了非整倍体,即增加了一条 Ph 染色体。在一项研究中发现在 67 例 CML 患者中有 33 人带有额外的 Ph 染色体和各种非整倍体,其中确定 28 人是 17 号染色体的长臂三体。在 58 例急性髓性白血瘤患者中有 21 人为 8 号染色体的非整倍体;16 人为 9 号染色体的非整倍体;10 人为 21 号染色体非整倍体。在另一项研究中发现 15 例肠癌患者中有 12 人带有异常染色体。它们常为 8, 13, 15, 17 或 21 染色体的三体。这些研究建立了染色体异常和癌的相关性,但具体的致癌机理尚不清楚。有可能影响到某些癌基因表达的异常所致。

四、非整倍体在植物育种中的应用

早在 1950 年基因工程(1972)尚未问世的前 20 年 E.R.Xears 就用经典的方法将叶锈病的抗性基因从伞花山羊草（*Aegilops umbellulata*）中转移到小麦中，育成了抗叶锈病的小麦(图 22－38)。现在的基因工程又称为体外重组，那么 Sears 的方法堪称"体内重组"了。

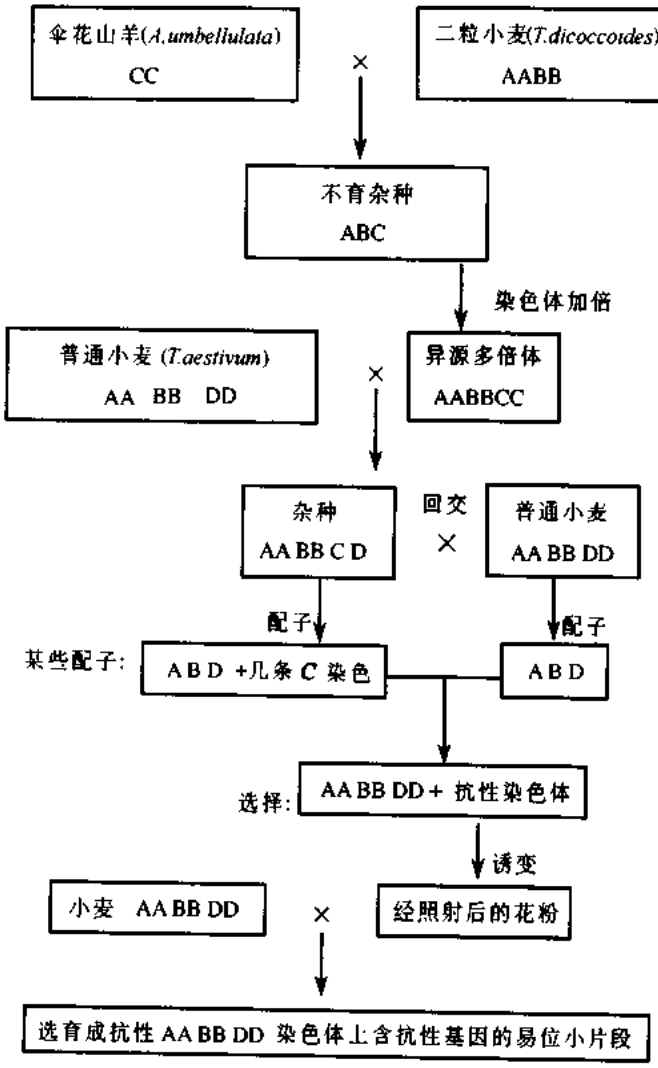


图 22－38　E.R.Sears 培育具有抗锈病小麦的技术路线。

叶锈病是对小麦危害很大的一种病害。而伞花山羊草这种禾本科的草却能抗叶锈病，也就是说它带有叶锈病的抗性基因。但这种野草不能和普通小麦进行杂交，而只能与其亲缘关系相近的二粒小麦(*Triticum dicoccoide*)杂交。Sears 就将二粒小麦作桥梁杂交(bridging cross)品系。伞花山羊草(2x = 14)其染色体与二粒小麦(2x = 28)并不同源，分别以 CC 和 AABB 表示，每个字母表示一套 7 条染色体。二者之间的种间杂种含有 3x = 21(ABC 是不育的)。经秋水仙素加倍后变成了异源多倍体(6x = 42, AA, BB, CC)再

将它和普通小麦(*T. aestivum*)(6x = 42, AA, BB, DD)杂交,它们二者之间三组染色体仅一组不同源,产生了杂合的 6x(AA,BB,CD)杂种,将些杂种再与普通小麦回交。杂种减数分裂对由于 CD 不能配对,因此随机地趋向一极,有的配子除了含有完整无缺 ABD 染色体以外还含有几条 C 染色体,它来自伞花山羊草,可能带有抗叶锈病基因。两种配子结合后产生多种类型的后代,Sears 选择其中具有抗性的后代,其必然带有抗性的 C 组染色体。但也可能带有不需要的 C 组染色体,再者即使是仅一条含有抗性基因的染色体也有不利的影响,因为其上还有很多我们不需要的基因。单条的染色体无论在减数分裂和有丝分裂中都只能趋向一极,有的配子就得不到它,在体细胞中也会形成嵌合体,为了解决这一连串的问题 Sears 采用射线照射,目的使含有抗性基因的染色体片段能易位到小麦的染色体上,这样既排除了不需要的 C 组其他染色体,也排除了大量与抗性基因连锁的多余的基因。而且抗性基因也被固定了,可谓一举三得。经照射诱变的花粉再授粉到去雄的普通小麦花上,产生的下一代中经反复选择终于获得了具有抗叶锈病的普通小麦,达到了将抗叶锈病基因转移到小麦中的目的。

第二十三章

重组和转座

基本原理

重组和转座的结果相同都引起了 DNA 的重排。转座是重组的特殊类型，它是由转座因子产生的特殊的行为。

同源重组是重组中的普遍形式，Holliday 模型，Meselson - Radding 模型较好地解释了同源重组的机制及基因转换和极化的现象。Szostak 的双链断裂起始重组模型较适合于真核生物减数分裂中的重组。

在原核同源重组中存在 8 个碱基组成的 *Chi* 位点，RecA、B、C、D 参与重组。RecBCD 复合体具有核酸酶、解旋酶和 ATPase 活性。在 *Chi* 位点产生单链 3'游离末端，RecA 可促进分枝移动。

在原核同源重组中 RuvA 和 RuvB 可促进异源双链的形成。RuvC 是一种内切酶可释放重组中间体。

λ 噬菌体的整合是通过 *att P* 位点(23bp)中和宿主 *att B* 位点(240bp)之间 15bp 同源区(核心区)交错切割，配对连接而无须经修复复制的过程。整合时需要 λ 编码的 Int 蛋白和宿主编码的 IHF 蛋白，切离时还须 λ 编码的另一种蛋白 Xis。

沙门细菌的相转变，Mu 噬菌体 G 片断的倒位及 P1 噬菌体 C 片断的倒位都需要短的反向重复序列，分别由 Hin，Gin 和 Cin 催化 IRR 和 IRL 之间的同源重组，产生侧位，Hin，Gin 和 Cin 是同源的。并和 Tn3 的 TnpR 蛋白同源。重组机制与 Tn3 共合体解离相似。转座子中 IR 之间的同源重组产生倒位；DR 之间的同源重组产生切离和缺失。

转座因子的类型如下(表 23 - 1)。转座因子的共同特点是：①两端有反向重复序列；②转座后靶位点重复正向重复；③都编码与转座有关的蛋白；④可以在基因组中移动。

表 23 - 1　转座子的类型

- 转座因子
  - 原核
    - IS - 两端有 IR，只编码转座酶
    - 类转座因子 - 结构同 IS，但不能独立存在，仅作为复合子的两端组件
    - 复合转座子 - 两端由 IS 或类 IS 构成，可编码抗抗菌素物质
    - TnA 转座子家族 - 两端为 IR，可编码转座酶、解离酶和抗性物质
  - 真核
    - 转座子
      - AC - Ds - 植物(玉米)中的激活 - 解离因子
      - P 因子 - 果蝇中父本因子，在 M♀ × P♂ 中导致杂种不育
    - 反转录转座子
      - 反转录病毒、RNA - > DNA - > 整合宿主靶 DNA
      - 病毒超家族：Ty、Copia、LINSL1 — (1)有长末端重复序列 (2)编码反转录酶或整合酶 (3)可含内含子
      - 非病毒超家族：SINSB1/Alu、假基因 — (1)无重复序列 (2)不编码转座子产物 (3)无内含子

转座子的转座机制分为三种类型：①复制型；②非复制型；③保守型。其共同特点是涉及靶位点的交错切割和修复性复制，主要是由 Shanprio 模型等加以解释。反转录转座子是由 RNA 介导的模板转换来实现的。

各种重组的特点和区别如表 23 - 2 所示。

遗传物质发生变异包括二个方面，一是突变，基因或染色体的结构发生改变；

表 23－2 各种重组的类型

		事件	同源性	蛋白和酶	特异位点	重组机制	行为和结果	复制
同源重组	真核	减数分裂	完全同源	内切酶、外切酶		Holliday 模型和 Meselson-Radding 模型，双链断裂模型	交互交换	修复性复制
	原核	转化	完全同源	RecA、B、C、D	Chi 位点（Slip）	Holliday模式	单向取代	修复性复制
							单向取代	修复性复制
		接合	完全同源	RuvABC			单向取代	修复性复制
		广泛性转导	完全同源	DNA pol				
位点特异重组	λ整合		区域同源	Int、IHF	attp核心酶（15bp）	交错切割连接Holliday连接	整合	无复制
	沙门氏菌相转变		区域同源	Hin 和 tnpR 同源	IP（34bp）	和Tn3解离相似	倒位	无复制
	Mu的G片段倒位		区域同源	Gin			倒位	无复制
	P1的C片段倒位		区域同源	Cin			倒位	无复制
	转座子之间或两端DR、IR重组		区域同源		DR、IR		DR→切离和缺失 IR→倒位	无复制
	反转录病毒整合			整合酶	靶	整合酶在LTR的3′端产生2b缺口在靶上交错切割连接填补缺口	整合	修复性复制
同源特异重组	酵母交配型转变		完全同源	Rid52、HO	特异Y/Z交界	交错切割、配对、降解，修复，产生Chi结构	单向取代	修复性复制
转座重组	复制型		无	转座酶	靶	转座子两端IR及靶的交错切割、连接、修复性复制和解离	单向插入	修复性复制
	非复制型		无	解离酶			单向插入	修复性复制
	保守型		无	DNA pol			单向插入	修复性复制
	I类内含子归巢		无	内切酶、外切酶、DNA pol、连接酶、解离酶		仅靶位点的交错切割与重组修复相似	单向插入	修复性复制
特殊重组	反转录转座子			反转录酶、内切酶	靶	RNA介导模板转换、交错切割	单向插入	修复性复制
	II类内含子归巢			反转录酶、内切酶	靶		单向插入	修复性复制
	免疫球蛋白的VJC连接		区域同源	RAG 1、2	特异位点（编码端）	平切→切单链→形成3′凸端→补平→加入碱基	剪接体细胞重组	修复性复制

二是**重组**，染色体的序列发生重排及新的组合。重组是遗传学的灵魂，没有重组就没有生物的进化；没有重组没有万紫千红的大千世界；没有重组也就没有现代的分子克隆技术。

重组的概念也有广义和狭义之分，广义的重组是指由于独立分配或交换在后代中

出现新的基因组合的过程。而狭义的重组仅仅是指基因的交换或重排而产生的重组。

重组的类型是很多的，我们可以将它分成以下几种主要的类型：

(1)同源重组(homologous recombination)或普遍性重组(generalized recombination)　反应涉及到大片段同源 DNA 序列之间的交换。在真核生物减数分减过程中发生的这种重组是一种**交互重组**的类型；在细菌中转化，接合和广泛性转导中所产生的重组子是一个**单向重组**，即仅使受体发生重组，供体并未发生改变。其主要特点是需要 RecA 蛋白的介入。

(2)位点特异性重组(site-specific recombination)　重组发生在特殊位点上，此位点含有短的同源序列，供重组蛋白识别。这种重组的特点是重组发生在特异位点；重组过程需蛋白的催化；重组过程涉及到交错切割。最为典型的是 λ 噬菌体在 *att* 位点整合到 *E.coli* 的基因组中。沙门氏菌鞭毛的相变位中 *H2* 基因和 *rh1* 基因上游一段 995bp 序列的倒位(见第 17 章)。Mu 噬菌体中 G 片断的倒位同样也涉及短的重复序列，Hin 蛋白和 Gin 蛋白的催化以及交错切割，所以也属此类。

(3)转座重组　转座的机制依赖 DNA 的交错剪切和复制，但不依赖于同源序列。转座涉及转座酶，解离酶和 DNA 聚合酶，共分为复制型、非复制及保守型三种类型。转座的过程中会形成**共合体**。两个转座因子之间的重组会引起缺失和倒位。

(4)模板选择(copy choice)性重组　适用于 RNA 病毒，在这种重组中，聚合酶以一个模板转换到另一个模板来合成 RNA，结果新合成的分子将含有两个不同亲体的遗传信息。内含子归巢也属于此种类型。

(5)特殊重组　在哺乳动物的体细胞中，免疫球蛋白成熟的 V.J.C(或 V.D.JC)区要进行 DNA 重组，这种重组涉及短的同源序列及特殊的剪切、修补、连接和插入碱基。

(6)同源特异重组　酶母交配型的转变(见第 18 章)和以上类型有相似之处，但又有明显的不同。如交配型转变同样涉及同源大片断的配对，但重组位点仅特异地在 Z/ Y 交界区，而且伴随着复制过程，切割时仅受体位点被切，这些特点很像复制型转座，但与转座重组的根本区别在于：① 位点特异；② 在结构上无反向重复，也不形成靶位点的正向重复；③ 不涉及转座酶和解离酶，因此作者认为应将其另列一个类型称为同源特异重组。

第一节　重组机制

重组是如此重要，那么重组理论毫无疑问也是十分重要的。1931 年著名的 SternC. 的果蝇实验和 Creighton H B，McClintock B 的玉米实验用细胞学方法直接证实了交换，似乎也表明重组就是染色体的断裂和重接，但问题并不是那么简单，科学家们先后建立很多重组模型，从 1909 年 Janssens 提出的交叉理论起已近九十年，至今这个问题还没有完全解决。证据也还不十分充足。现在我们来看看前人怎样向前探索。

一、交叉理论

最早的重组理论是在 1909 年(此时 Morgan 的连锁定律尚未问世)比利时的细胞学家 Janssens 在研究蝾螈和直翅目昆虫的减数分裂时，观察到二价体的交叉，提出的交叉型学说(chiasmatype hypothesis)。认为每次交叉都表明父、母本一条染色单体接触、

断裂和重接，形成一个新的组合，其他两条染色单体仍保持完整状态。因此每个交叉是重组行为可见的表现形式。交叉理论虽然能得到遗传学家的支持，但却遭到细胞学家的反对，因为找不出理论中有关断裂和重接的任何证据。从显微照片上来看确实很难想象存在断裂和融合的现象。直到 1928 年 Belling 提出交换发生在减数裂较早时期，此时同源染色体紧密相连，因此在双线期观察到的交叉并不表明交换的过程，而是交换的直接结果，这一解释结束了近十年的争论。但有一个问题仍使人们感到困惑，那就是交叉点是在变化的，在接近中期 I 时明显减少了，而交换一旦发生数目应是固定的，因此仍然怀疑交叉型学说。1931 英国的植物细胞学家 Darlngton 发现交叉数目减少是交叉端化的结果，此时才排除了异议（图 23－1）。

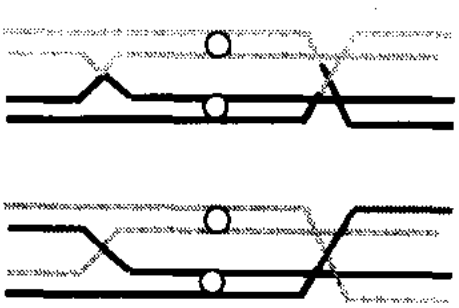


图 23－1　交叉型学说(a)与传统理论的比较。

1978 年 C.Tease 和 G.H.Jones 采用姊妹染色体单体差别染色体(sister chromatid differential staining ，SCD)来直接检测交叉和重组的关系。SCD 法可将姊妹染色体单体一条染上颜色呈现深紫色，显示为"暗"的。另一染不上吉姆萨，显示为"明"的。他们用 SCD 法来处理减数分裂的染色体，结果发现明－暗转换发生在交叉处(图 23－2)。这就澄清了前面留下来的问题。实验表明交叉正是发生交换的位点。

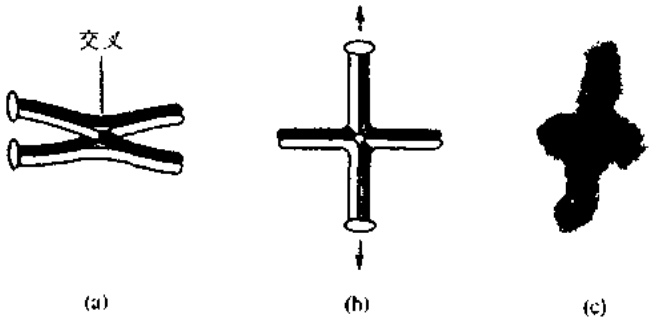


图 23－2　在减数分裂中非姊妹染色单体深色链和浅色链之间发生交换。(a)表明交叉；(b)当着丝粒附着到纺锤丝上拉向两极时，形成了中央带有交叉的十字形结构；(c)十字结构的照片。(转引自 Griffiths A.J.F. *et al*.: An Intiroduction to Genetic Analysis.5th ed.. 1993,Fig.19－2)

二、断裂和重接模型

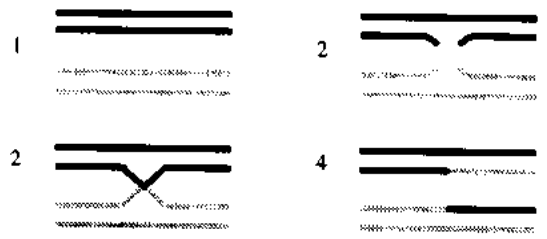
1937 年 Darlington 仔细地观察研究了减数分裂过程，提出了重组的断裂和重接的模型，他认为在减数分裂中一对同源染色体相互分离就像将绳子的两股分开一样，会产生扭曲，为了消除张力，只有当两姊妹染色单体在对应点发生断裂才能使张力达到平衡，然后非姊妹染色体单体的"断头"相互重接，产生了重组。根据这个模型重组应发生在四线期，因重组是一个相互的过程，所以重组的产物也应当是对称的。但 1930 年德国的遗传学家 Winkler 已发现基因转变(gene conversion)的现象，即链孢霉重组的产物并不都是按 4：4 分离，也会出现 5：3 和 6：2 的分离比，断裂重接模型对此无法解

释。约在 1955 年人们已认识到基因是 DNA 分子的一部分，当遗传学家发现基因的转换说明基因内重组时更加怀疑断裂重接模型的正确性。特别是考虑到此模型要求两条非姊妹染色单体必须要在相同位点的核苷酸之间发生断裂，否则重组的产物其长度就会改变。张力引起的染色单体断裂能达到如此之精确也是难以置信的。于是人们便热衷于寻求一种新的模型，要求它能很好地解释基因的转换，于是模板选择学说（copy choice）受到了推崇。当时虽然还不清楚在噬菌体中 DNA 半保留复制的过程是怎样进行的，但在 20 世纪 50 年代大部分分子遗传学家都倾向于模板选择说。

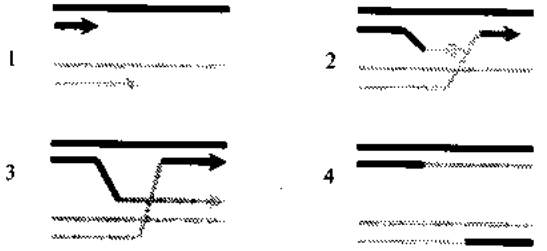
三、模板选择学说

模板选择学说的兴衰是遗传学史中有趣的一章。这一学说一开始是由 Belling J 首先提出的，1933 年他又撤回了这一假设。1948 年 Hershey 发现在噬菌体的杂交中产生的重组子有时是非对称的，为了解释这个现象，他接受了 Sturtevant AH 的建议提出了在噬菌体中重组可能不是遗传结构的断裂和重接，而是复制时改变了模板所致。即两条染色单体作为复制的模板，新的染色体各以一个单体模板进行复制，在复制过程中相应交换模板，从而造成重组（图 23－3）。模板选择模型能够很好地解释基因转变的现象，因此在 20 世纪 50 年代兴盛起来。这是由于当时半保留复制等有关机制尚不清楚，若根据现在已知的理论就不难发现，此模板选择学说是不能成立的。①模板选择假设提出一条姊妹染色体单体以另一条单体为模板进行复制这本身就违反半保留复制的原则，即以一条双链 DNA 为模板合成另一条新的 DNA 双链，此属于全保留复制。②在真菌中可以观察到染色体单体的三线和四线交换，而模板选择模型仅限于二线交换，此也难以成立。染色体的配对、交换应在细胞周期的分裂期（M），而复制是在 S 期，因此重组不可能是和复制同时发生。

更为有力的否定证据是 1951 年 Meselson M 和 Weigle J 所做的实验。他们用两个



(a) 断裂重接模型



(b) 选择复制模型

图 23－3　断裂重接模型和选择复制模型的比较。

品系的 λ 感染 *E.coli* 一个品系在染色体的一端有遗传标记 *C* 和 *mi*，这个品系用 ^{13}C 和 ^{15}N 进行重组标记；另一个品系的相应座位带有 C^+ 和 mi^+ 遗传标记。并以正常的 ^{12}C 和 ^{14}N 作为轻标记(图 23－4a)。复感染以后在轻标记(正常 ^{12}C 和 ^{14}N)培养基中培养宿主 *E.coli*，直到裂解。

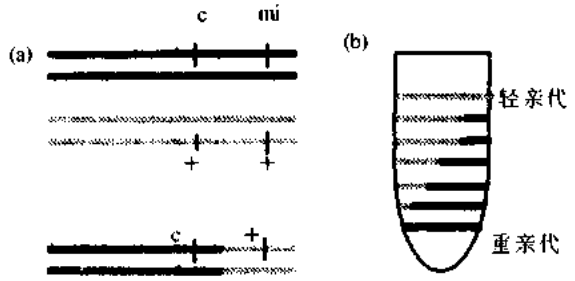


图 23－4　轻、重标记 λ 噬菌体的重组实验。(仿 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 19－1)

后代噬者体从裂解细胞中释放出来，收集后进行 CsCl 密度梯度离心，结果获得了很宽的噬菌体带，表明它们的 DNA 从重标记的亲代到轻标记的亲代从下自上分布。中间还有很多不同密度的带（图 23－4b），有趣的是某些重组的噬菌体带分布在靠近重标记亲代的区域。经分离鉴定它们的基因型是 $C\ mi^+$，它们必定是通过 *C* 和 *mi* 两个标记基因之间发生交换而产生的(图 23－4C)，因为 mi^+ 位于轻标记链的末端，所以重组后代的染色体大部分是重标记，故靠近底部的重标记亲代。若重组是复制的结果，复制的后代都应是轻标记，不可能位于底部，因此重组仍是断裂重接的结果而不可能是模板选择而产生。

第二节　Holliday 模型

前面已经提到早到在 1930 年 Winkler 把真菌中不规则分离现象解释为减数分裂过程中同源染色体联会时，一个基因使相对位置上的基因发生相应的变化所致，因而称**基因转变**（conversion）。Lindegren CC(1949)报道了有规律的异常分离现象：在酵母（*S. cerevisiae*）不同交配型 A×a 的杂交中，有些子囊所含的孢子为(3A＋1a)，而不是预期的(2A＋2a)。他把这一现象也称为基因转变。直到 50 年代中期基因转变这一现象才得到进一步的认识。当时 MitchellM B 等将不同的 B6 突变型链孢霉进行杂交，结果在子囊中拟等位基因 *pdx* 以正常的 4∶4 分离，而与其紧密连锁的拟等位基因 *pdxp* 却出现了 6∶2 的异常分离(图 23－5)。表明异常分离不是由于整个染色体的异常行为，而是由于特定位点的“基因转变”所致，即属于基因内重组。

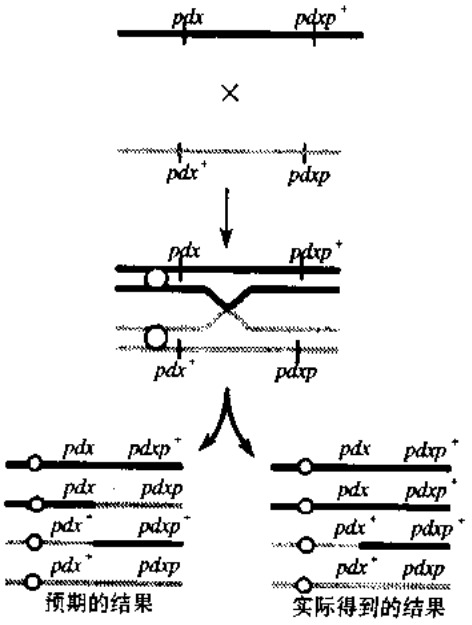


图 23－5　链孢霉中的基因转换。

Olive，El-Ani 和 Kitani 等在粪壳菌(*Sodaria fimicola*)中也发现了异常分离。在野生型 g^+(子囊孢子为黑色)和 g^-(子囊孢子为灰色)杂交中，观察了 20 万个子囊，其中大部分为 4：4 的正常分离，少数为异常分离：0.06%为 5：3 分离；0.05%为 6：2 分离，0.008%为异常的 4：4 分离(图 23－6)。如果染色体中每一条双链都是纯合互补，减数分裂离生的四个子囊孢子再经一次有丝分裂产生的 8 个子囊不论是否发生重组都应当是相同的基因型成对排列，所出现的“+ + + — + — — —”以及“+ + + + + — — —”一定是有异源双链的存在。这种异常分离是有丝分裂的产物，故称为**减数后分离**(postmeiotic segregation)。这些资料另一显著特点是异常分离的子囊中约有30%的子囊在 g 位点的两侧标记之间发生了交换。其中52 例中有 48 例既有 A 位点的重组，又有g 位点

图 23－6　粪壳菌(*Sordaria fimicola*)异常分离的子囊。

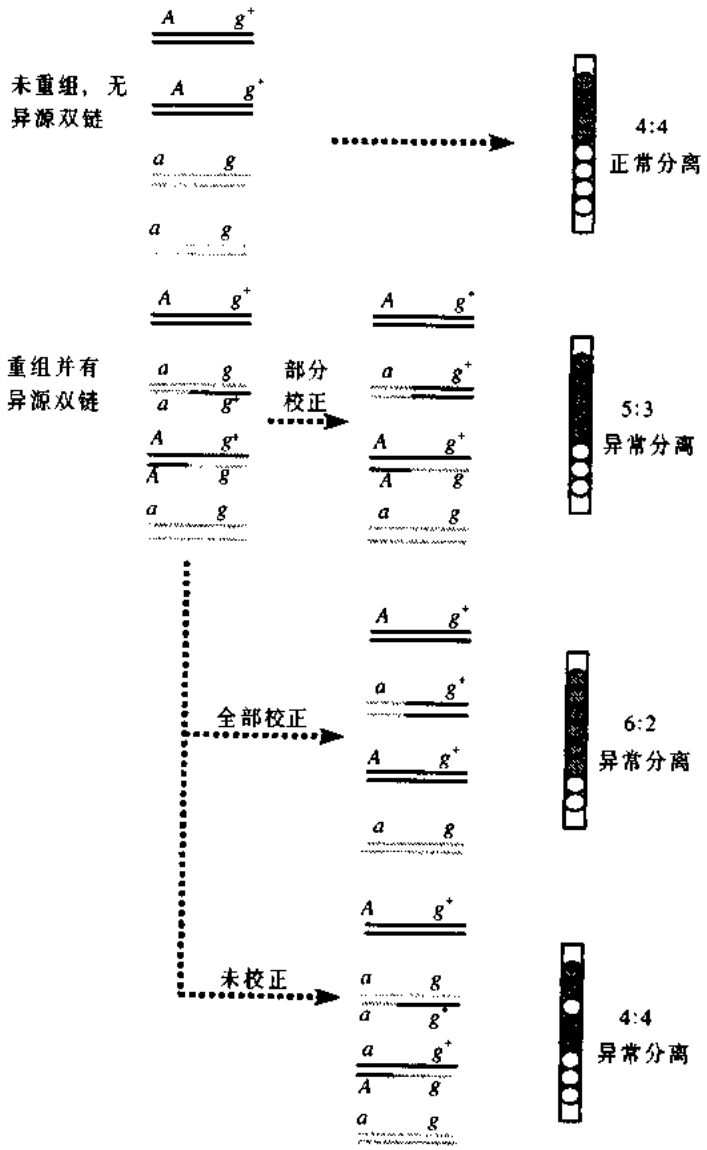


图 23－7　基因转变和减数后分离。

的基因转变(图23－7),表明基因转变与重组有着密切的关系。

基因的转变是基因内部不同位点之间的重组,不同位点可以单独发生转变,也可同时发生转变,同时发生的基因转变称为**共转变**(coconversion)。基因的转变常涉及 1kb 的范围,更长的距离为基因间的重组。当两个突变位点距离越近,共转变的频率越高,距离较远时常发生单个突变位点的转变。同一基因内部的各个突变位点的转变频率常从基因的一端向另一端逐步递减,具有这种现象的一段 DNA 称为**极化子**(polaron),一个极化子相当于一个基因(Rizet 等,1961)。

解释基因转变的模型都必须解释异源双链的形成以及基因转变往往伴随着两侧

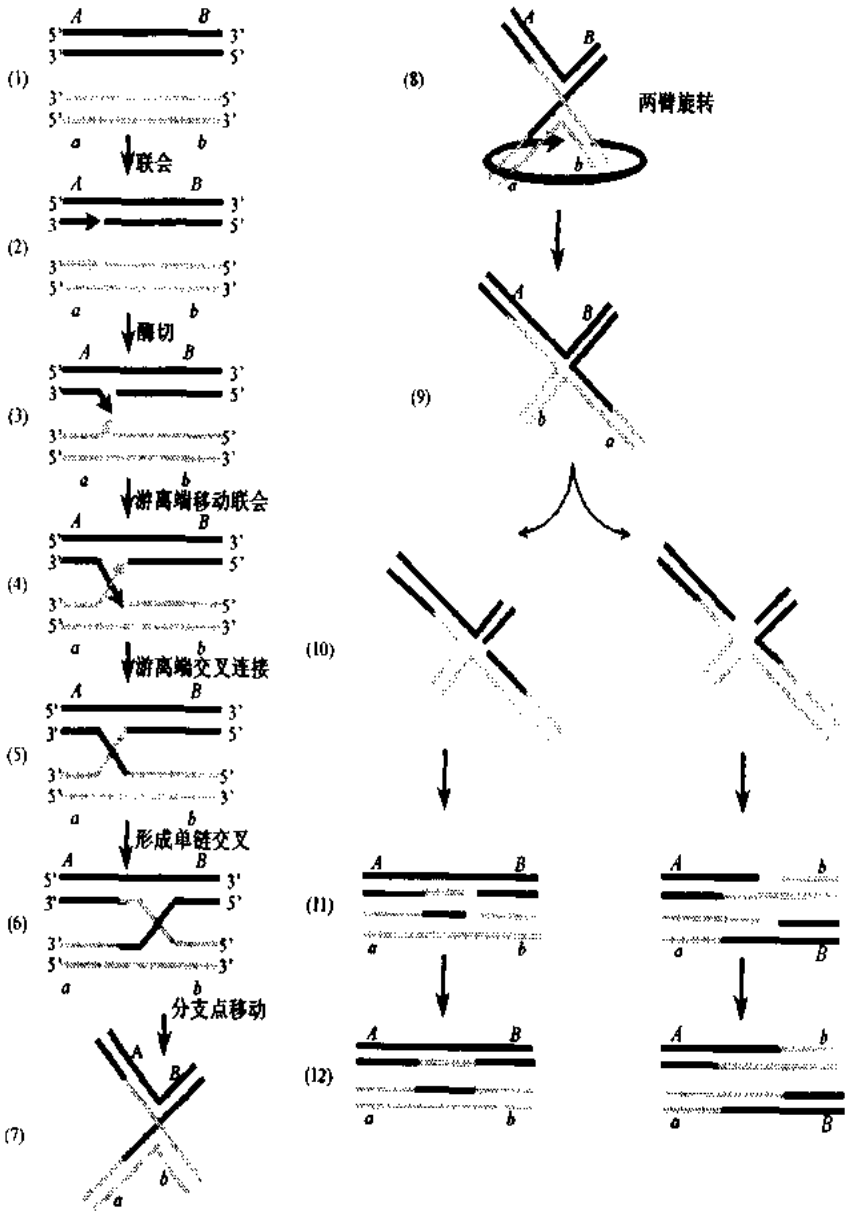


图 23－8　Holliday 重组模型。(1)同源染色体联会(图中只图出四条染色单体中的二条非姐妹染色单体;(2)两条方向相同的单链被切断;(3)游离端的氢链断裂,离开互补单链;(4)游离端交叉地和断裂单链连接;(5)连接成半交叉,称为 Holliday 结构;(6)(7)分支点移动呈"十"字型;(8)"十字型"的两臂旋转 180°,成为(9)中空的十字结构(10)两种不同的解离方式产生(11);经修补连接后形成(12),中部有异源 DNA 环链。两侧的标记有一半发生重组。

基因重组这一现象。1964 年美国学者 Robin Holliday 提出了著名的 Holiday 模型（图 23－8），很好地解释了高等的真核的重组和真菌中发现的基因转变的现象。以后其他的一些学者又在此模型的基础上提出了修改的模型，特别是 Meselson-Ralddding 重组模型（图 23－9）影响较大。近代 H. Potter 和 D. Dressler 在噬菌体中的拍摄到 Hollidy 结构的电镜照片（图 23－10），为 Holliday 模型提供了一个有力的证据。以上两种模型可以解释下面的几种现象：

（1）染色单体转变和半染色单体转变。

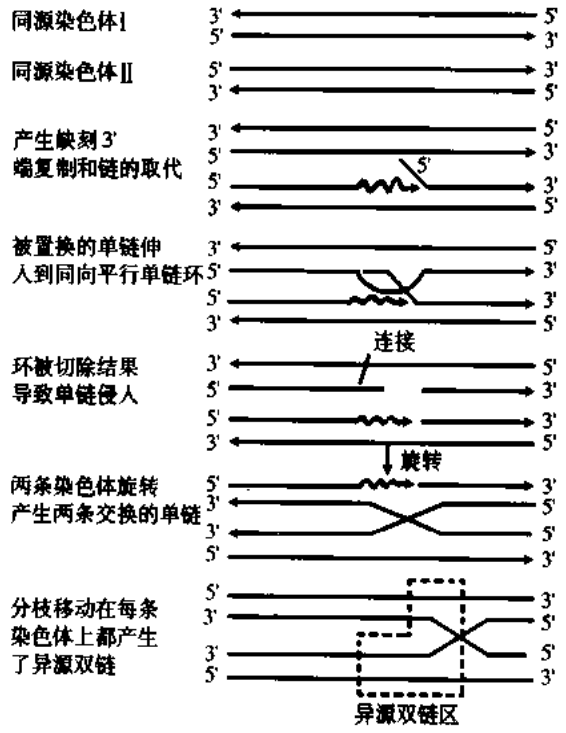


图 23－9　Meselson－Radding 的重组模型。

（引自 Griffiths, A. I. F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 19－11）

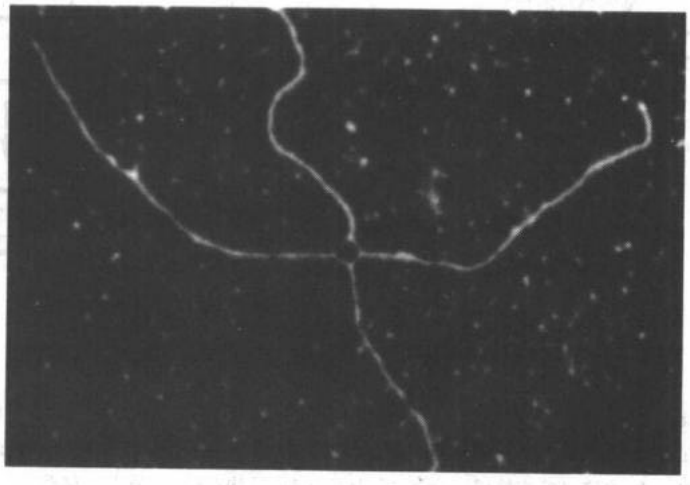


图 23－10　DNA 在分支点带有一些单链的十字形 Holliday 中间体的电镜照片。

（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 11－25）

在图 23－7 中，表明在真菌的杂交中有时会产生一些异常分离，如出现 6：2 或 2：6 的比例，看来似乎在减数分裂中有一条整个的染色体发生转变，这一过程就称为**染色单体转变**（chromatid conversion）；另外也有 5：3 或 3：5 的异常分离，看起来似乎是半染色体发生了转变，故称为**半染色单体转变**（half-chromatid conversion）。

以上的异常比例可以解释为在交换的过程中产生了异源的链和错配（图 23－8，9），由于不同的校正而产生了三种不同的异常比例（表 23－3）。

表 23－3　交换过程中异源链的错配和不同的校正

减数分裂开始阶段的 DNA 链	产生异源双链时的 DNA 链	不校正	一条异源双链的 m 校正为 +	二条异源双链的 m 都校正为 +
m	m	3m	3m	2m
m	m			
m	m			6 +
m	+	+	5 +	
+	m	m		
+	+	3 +		
+	+			
+	+			
一对同源染色体	4 条染色单体	异常 4:4	异常 5:3	6:2

（2）极性（polarity）

在基因中可以精确地制作不同突变位点的遗传图。我们可以将基因中不同位点的转变频率进行比较就不难发现基因一端的位点其转变频率高于另一端，即转变的频率沿着基因呈现出一种极性。这是由于交换产生的错配只发生在异源双链 DNA 这部分，而这一部分只发生在 Holliday 结构的断裂和分枝点之间，离断裂点越远的基因座位离分枝点越远，而不处于异源双链的部分。在分枝点解离时，外侧区域仍然形成同源双链（图 23－11）。

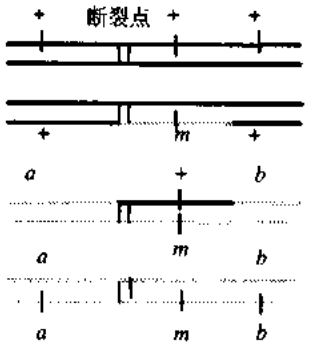


图 23－11　基因 b 远离断裂点，在异源双链区之外，发生转变的概率较低。

第三节　双链断裂起始重组

一、双链断裂起始重组模型

Szostak J W 等（1983）提出了双链断裂起始重组的模型（图 23－12）。他们提出重组是由一种内切酶起始的，内切酶切开一对 DNA 双链中的一条受体链。通过外切酶的“扩创”，将切口扩展成一个裂缺（gap），外切酶“咬”住断裂链的一端，另一端游离侵入到另一条供体双链中，异源双链的形成同时也产生了 D 环，取代了供体双链中的一条链。用游离 3′端为引物，通过修复合成使 D 环延伸。

最终 D 环扩大到足以和受体染色体单体的缺口长度一致。当单链延伸达到缺口远端时，互补单链序列退火，现在在缺口的任一侧都有异源双链 DNA，缺口本身被 D

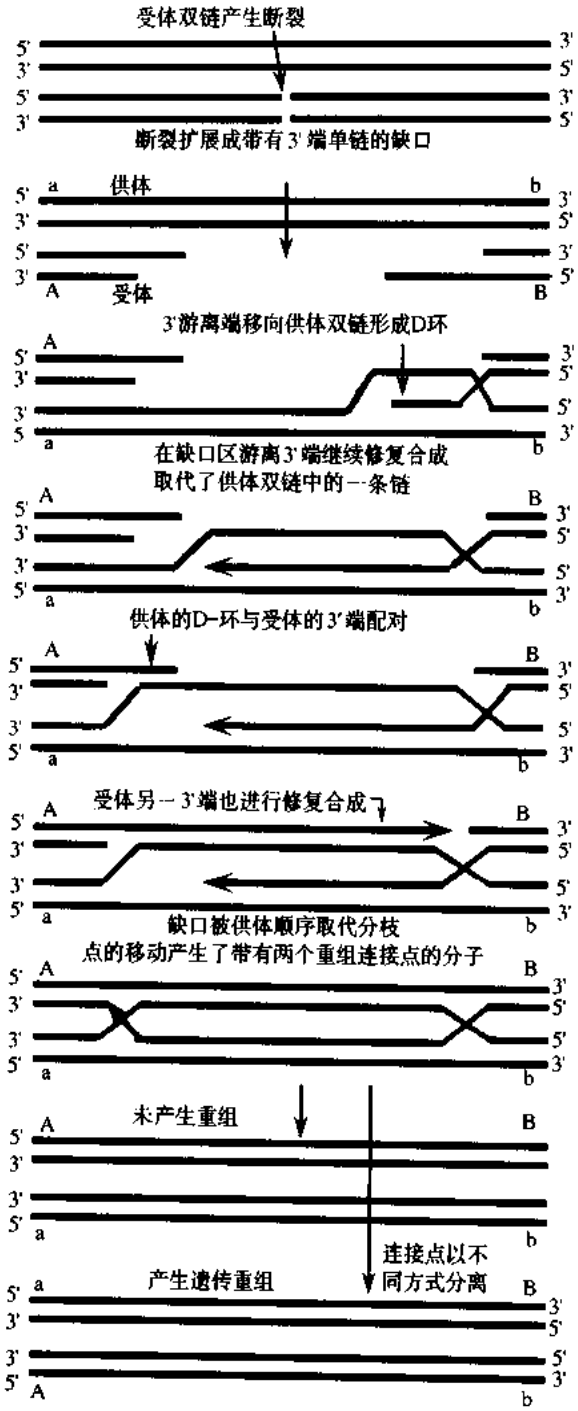


图 23－12　Szostak, J. W. 的双链断裂起始重组模型。

(参考 B. Lewin: Genes Ⅳ, 1990, Fig. 32－4; Genes Ⅵ, 1997, Fig. 17－5)

环的单链所取代。缺口左侧的 3′作为引物进行修复合成，使缺口区域被恢复。分枝点的迁移使这个结构变成了带有两个重组连接点的分子（图 23－12）。连接点必须通过剪切而被解离。解离的方式有两种(图 23－8)。

若两个连接点都以相同的方式解离，那么原来未交换的分子将被释放出来，每个都带有一段遗传信息发生改变的区域，此就是交换(exchange)的痕迹。若两个连接点是以不同的方式被解离，那么将产生遗传重组(genetic crossover)。

按双链断裂模型在和交换(exchange)有关的末端区域可形成异源双链,这个区域和供体分子中的缺口相一致。

二、重组和联会复合体之间的关系

重组中的主要矛盾是从未将亲代染色体与DNA的重组联系起来。染色体进入减数分裂已形成了姊妹染色单体。染色单体的数目是可以观察到的。同源染色体配对时形成**联会复合体**(synaptonemal complex, SC)。多年来有很多的推测,认为SC可能和重组的初始阶段有关,它对于DNA的变换(exchange)可能是必要的。比较新的观点认为SC是重组的结果,而不是重组的原因。无论是因还是果,我们还不清楚,SC的结构与DNA分子间的连接有什么相关性。

开始联会时每条染色体(含一对姐妹染色单体)固缩为轴向分子(axial element),它明显带有蛋白质性质。然后染色体的轴向分子平行排列,形成SC。SC由三部分构成(图23-13):①轴向分子现称为侧翼成分(lateral elements);②中央成分(central element),位于2个侧体中间;③染色单体。

在此阶段每条染色体是以染色单体被侧翼成分结合成束而存在。两个侧翼成分都在中央成分的两侧,三联体呈链状平行排列于一个平面上,延着轴弯曲和扭转。同源染色体的距离大于200nm,而DNA的二聚体的距离是1.2nm。这样就出现了一个问题,既然SC是同源染色体排列而成,相距这么远如何使同源DNA分子接触连接呢?

在SC两侧之间的连接仅在真菌和昆虫的球面或柱状结构中见得到。它们位于复合体的交叉处,被称为"节结"(nodes)或**重组节**(recombination nodules)。它们以相同的频率发生并以"交叉"的形式存在。它们的名字反映了人们认为此处可能就是重组位点。

重组和SC形成之间的相关性将会被建立。最近的工作表明,在果蝇和酵母中染色体不能配对的突变也会阻止重组。已有几个系统能在分子水平和细胞学水平来比较与重组的关系。近年来分析了酵母(*S. cervisiae*)的减数分裂,如图(23-14)所示。

这对于同源重组和位点特异性重组中双链断裂起始重组是一个很好的证据。双链

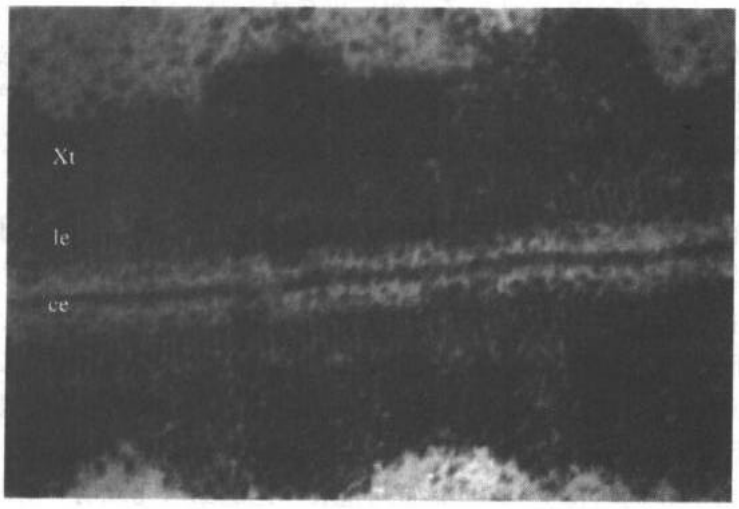


图23-13 Neotellia的联会复合体。照片表明染色体配对时其同源DNA并未并列地排在一起。Xt:外侧的染色单体;le:侧体,直径约50nm;ce:中体,直径约18nm。(转引自B.Lewin:Genes, Ⅵ,1997,Fig.17-16)

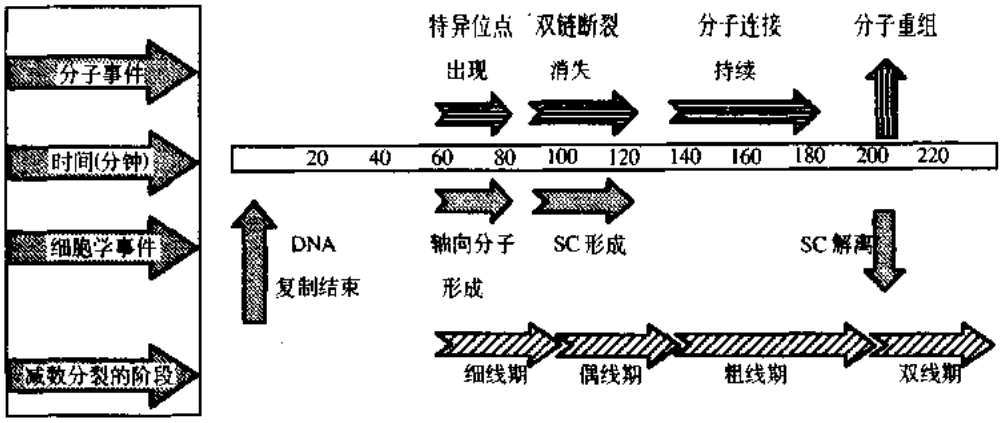


图 23－14　在酵母中重组的分子事件和细胞学事件的相关性。(B.Lewin:Genes Ⅵ,1997,Fig.17－7)

断裂开始与交配型转变有关系(见第 19 章),它还涉及到复制。减数分裂早期在特异位点也发生双链断裂,它提供了重组热点。在这个位点的邻近地区重组的频率很高。在热点的一侧或两侧重组频率以一个梯度倾斜。经鉴别重组就是在热点起始的。这种倾斜的梯度反映了重组可能在扩散。

我们现在可以从分子水平来解释这些现象。由双链断裂而产生的平端迅速在两侧变成为单链末端(图 23－12)。在突变的酵母中(*rad50*)阻碍了平端变成突出的单链而不能重组。这表明双链断裂对于重组来说是必须的。而梯度是取决于倾斜的概率,即当距双链断裂位点的距离增加时产生了单链区域。

双链断裂从出现到消失每个周期 60 多分钟。分子连接推测可能是重组的过度形式,双链断裂后很快就消失了。重组事件的次序是双链断裂、配对反应和形成重组结构,在同一个染色体位点上连续地进行。

当轴向分子形成时双链断裂出现,染色体配对变成 SC 时,双链断裂又消失了。这些事件发生时间的相关性表明 SC 的形成是双链断裂诱发重组起始的结果。这一想法是由于 *rad50* 实变体不能使轴向分子变成 SC 而提出的假设。这一观点否定了减数分裂的传统观点:认为 SC 的存在是分子重组前染色体配对所需的。

决定在联会阶段是否发生重组是困难的,因为重组是通过减数分裂完成后重组体的存在来评估的。在酵母中是通过含有特殊限制性位点的 DNA 分子产物来直接评价的。它能表明重组出现在粗线期的末尾。很清楚的是重组事件的完成在 SC 形成之后。

联会复合体的形成是在起始重组的链断裂之后,一直持续到重组分子的形成。SC 的出现并不是为了重组的需要,因为有些突变体失去了产生正常 SC 的能力,但却能产生重组体。不能进行重组的突变也不能形成 SC。这表明 SC 的形成是重组的结果,染色体的配对对于减数分裂的前期阶段是必要的。一种酵母的突变(*zip1*)在 SC 形成中失去了重组的干涉(一个重组事件能抑制邻近地区另一次重组的发生)。这表明有可能 SC 的形成能在重组位点开始,延着染色体传播,从而抑制了重组的进一步发生。在这种情况下,重组事件并不同时发生,SC 的形成可能在重组起始后完成前一阶段。这就意味着 SC 对重组的抑制又是一件和早期观点相吻合的事。

从分子水平上我们可以假设,同源染色体联会之前,双链断裂可以产生单链末端,

此凸出的末端立即寻找互补的序列，结果与同源染色体产生了连接，它们靠得很近，起始了 SC 的形成。还不清楚在粗线期 DNA 重组前发生了什么变化，在这一段时间里可能进行双链断裂后的一些重组步骤，如单链的置换(exchange)，DNA 的合成和解离。

在减数分裂的下一阶段（双线期）染色体脱离了 SC，然后在染色体上出现了可以看到的交叉点，它表明发生了遗传物质的交换，但"交叉"的分子特征尚不清楚，也可能是完成交换后留下残迹，也可能是遗传交换尚未解离。在减数分裂的后期交叉移向染色体的两端。这些特点表明存在一些重组事件的遗迹，而不是重组的中间过程。

重组事件发生在减数分裂染色体的各个点上，但我们尚不能将重组的事件和这些点的结构联系起来，这些结构就是一些已发现的重组节结和交叉。涉及酵母重组的一些蛋白可定位在不连续的位点上，这些蛋白使我们对这些不连续结构形成的分子基础，有了一定的认识。它们包括和细菌错配修复的蛋白相关的 MSH4 (mismatch-repair)，Dmc1 和 Rad51(*E.coli* RecA 蛋白的同源蛋白)。这些蛋白在重组中的确切作用尚待确定。

重组事件受普遍性调控，只有少数的相互作用才能发展成为交换，典型的方式是每一次同源染色体的配对，仅获得 1～2 次交换。一次同源配对不发生交换的概率是很低的（$<0.1\%$），这一过程可能是单个的交换控制(crossover control)的结果，因此交换的非随机性常在一定的突变体中被破坏。

第四节　重组和酶

要分析 DNA 分子之间序列的交换，我们必须要回到细菌系统中。识别反应涉及 DNA 分子重组机制和限制区域的一部分，而不是整个染色体。但分子事件的一般次序是相同的：① 由断裂分子产生的单链与非姊妹染色单体的双链相互作用；② 配对区的延伸；③外切酶解离配对的双链。现在已经知道以上每一步都需要酶，可能对重组来说只需要其中某些成分。

在细菌中用于重组的酶已通过 *rec*$^-$ 突变而被鉴别出来。*rec*$^-$ 突变体的表型是不能进行重组。现已鉴别出 10～20 个座位。

在细菌中 DNA 双链的大小并不常发生改变。在原核生物中有多种途径都可导致重组。有时 DNA 可能会产生游离的 3′末端，如细菌的接合(见 8 章)；通过放射线照射产生损伤时可以形成单链缺口(见 21 章)；噬菌体的滚环复制也可产生单链的尾巴。然而在真核减数分裂中重组必定要涉及两个两倍性分子，而且必须产生单链区域和 3′末端。

已发现了一种产生合适末端的机制，它是重组热点存在的结果。在 λ 噬菌体中发现一种突变型，称为 *chi*(十字型)，它由于单个碱基的改变产生了刺激重组的位点。这个位点又把我们注意引导到另一些和重组有关的蛋白质上。

这一位点是由 8 个碱基组成的非对称序列。

5′GCTGGTGG 3′

3′CGACCACC 5′

这个 *chi* 序列在自然的 *E.coli* DNA 中大约每隔 5～10kb 就有一个拷贝。野生型 λDNA 无此序列，而其他遗传因子中也无此序列，表明它不是重组所必须的。

chi 序列促进了其邻近区域相距 10kb 的范围发生重组。*chi* 位点能被双链断裂所激活，断裂在其右侧几 kb 的地方。这种方向的依赖性表明重组的装置必须和 DNA 的断裂端相连接，然后才能延着一个方向在双链上移动。

chi 位点是 *recBCD* 编码的一种酶的靶序列。这个复合体有几种活性:①强的核酸酶,可降解 DNA,初始将它鉴定为核酸外切酶 V;②它有解旋酶活性,在 SSB 存在时它可以解开双链;③它有 ATPase 活性。在重组中的作用是提供带有游离末端的单链区。

图 23 - 15 表明 RecBCD 在具有 *chi* 位点的 DNA 上如何进行反应。当 Rec BCD 结合在 DNA *chi* 位点的右侧时，它沿着 DNA 移动，使 DNA 解旋，并降解带有 3′端的单链。当到达 *chi* 位点时它停顿下来,并切开上面的一条单链 DNA,切点在 *chi* 3′端离以 *chi* 4 - 6bp 处。*chi* 位点的识别导致 RecD 亚基的解离或失活,结果使复合物 RecBC 失去了核酸酶的活性,但继续发挥其解旋酶的作用。

RecBCD 介导的解旋和剪切可产生 3′游离末端，用来起始异源双链连接的的形成。当 RecBCD 在 *chi* 位点剪切时 RecA 酶能使带有 3′末端的单链释放出来,并利用它和同源双链序列起反应,从而产生连接的分子。

RecA 有两个完全不同类型的活性：①在 SOS 反应中它具有蛋白酶活性（见第 21 章);②能启动 DNA 单链和与之互补的双链分子进行碱基配对。

RecA 的后一种活性需要单链 DNA 和 ATP; RecA 的活性与体内重组之间的相关性尚不十分清楚,但和几种类型的重组机制的反应有关。

RecA 操作 DNA 的活性之一能使单链 DNA 取代双链分子中同源的链,此反应称为**单链取代**(single-strand uptake)或**单链同化**(single-strand assimilation)这种取代反应能发生在各种形状的 DNA 分子之间（图 23 - 15)，但一般要具备三个条件：① 其中一个

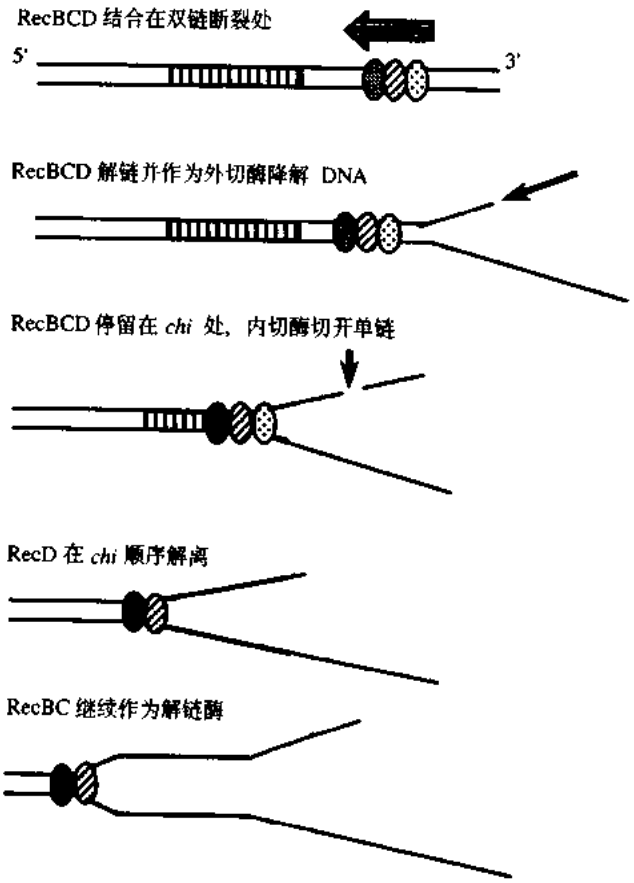


图 23 - 15　RecBCD 核酸酶从一侧接近 *chi* 顺序，当它前进时降解 DNA。在 *chi* 位点它作为内切酶剪切单链,释放 RecD,只保留了解链酶活性。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 17 - 8)

DNA 分子必须要有一个单链区域;②其中一个 DNA 分子必须要有一个游离的 3′末端;③此单链区域和 3′末端必须和另一个 DNA 分子有互补区。

反应如图 23-16 所示，当一个线状单链分子侵入到双链分子中，并取代一条同源链。这个反应很容易由环状的供体或受体所产生，反应的过程是单链延着 5′→3′方向置换和取代双链中的一条链。此反应和交换有关,在反应中交换链要有游离的 3′末端。

单链取代可能和重组的起始有关。在各种重组模型中都要求在单链和另一条双链交换中要有一个中间体(图 23-8,9,12)。RecA 可能催化反应的这一阶段。

RecA 促进分枝转移的机制是根据它能和单链 DNA 或双链 DNA 结合聚集成长丝状而提出的。这个长丝状结构呈螺旋状，其深沟中含有 DNA 分子，每个螺旋有 6 个 RecA 单体，因此每个 RecA 单体结合 3 对核苷，结果使 DNA 的形态拉长了，比 B-DNA 伸长了 1.5 倍,使每个螺旋有 18.6 核苷酸对。当 DNA 和 RecA 结合时,RecA 位于 DNA 的小沟中,而空着的大沟可能留着与另一个 DNA 分子反应,(图 23-17)。

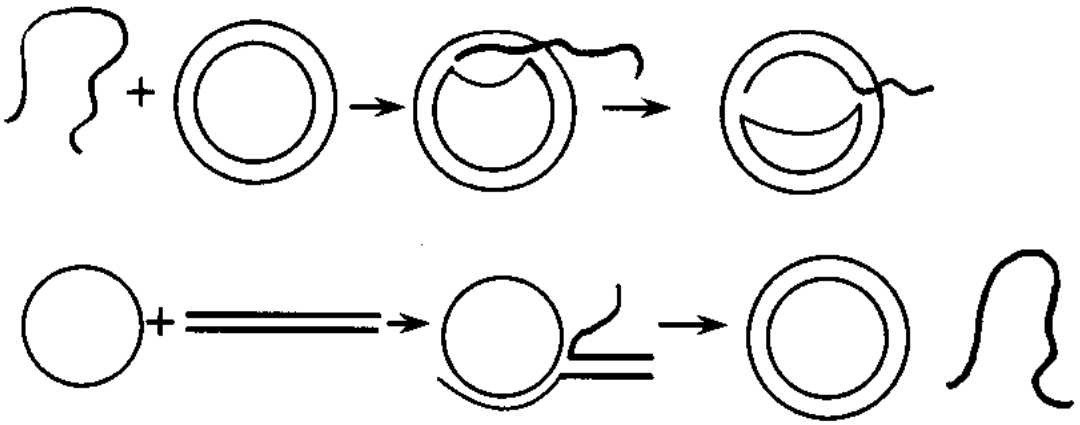


图 23-16 RecA 蛋白使单链 DNA 取代双链 DNA 中的同源部分。(参考 B.Lewin: Genes Ⅵ, 1997, Fig. 17-9)

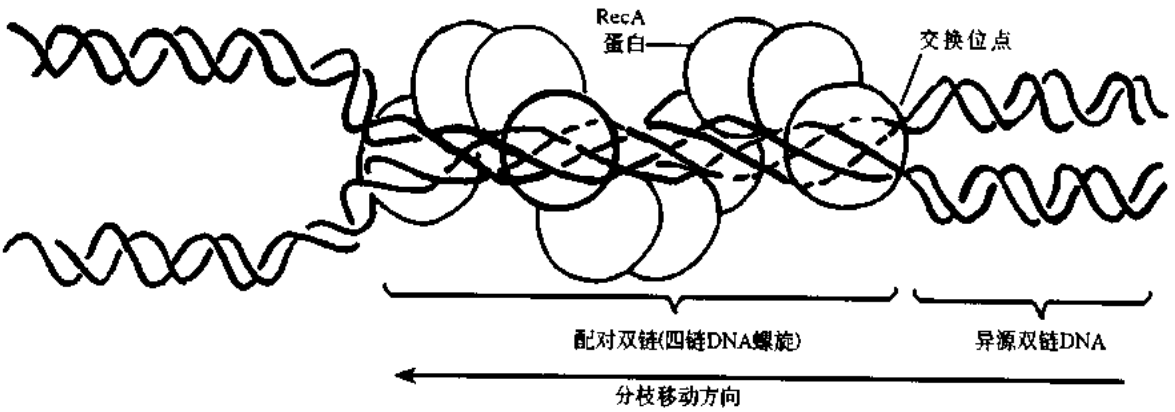


图 23-17 RecA 促进分枝移动的机制。(仿 Watson, J.D. *et al.*: Molecular Biology of the Gene, 4th ed., 1987, Fig. 11-15)

在这个长丝中两个 DNA 分子之间相互作用。当一个单链侵入到双链中时，第一步是 RecA 结合于单链，并使其纳入"细丝"中。然后以形成三链的方法合并成四链螺旋。在这个系统中联合进行遗传物质的物理交换，这样在单链缺乏游离末端，不能发生交换时也能发生配对反应。

游离 3′端对于单链交换是必要的，这个反应发生在"细丝"中。RecA 将它结合到双链中，并与其末端起作用。在反应时水解大量的 ATP。ATP 可使 RecA 的构象发生变构。当和 ATP 结合时，RecA 的 DNA 结合位点对 DNA 是高亲和性的，这对于结合 DNA 和配对反应都是必要的。ATP 的水解使这个结合位点变成低亲和性的，这对释放异源的双链 DNA 也是必要的。

我们可以将 RecA 催化单链和双链 DNA 之间的反应分为三个时期：①缓慢的联合前时期。在此期 RecA 聚集在单链 DNA 上；②快速的配对反应。在单链 DNA 和双链中其互补链之间产生异源双链连接。③缓慢的取代反应。取代双链中的一条链，产生一个长的异源双链区。

SSB 蛋白的存在促进了以上反应，其作用是使底物失去二级结构。现在还不清楚 SSB 和 RecA 蛋白二者是否能作用同一条 DNA 链，RecA 和 SSB 一样需要一定的化学量，表明在使单链侵入时，它的作用和协同结合 DNA 形成细丝结构有关。

当单链分子和双链 DNA 反应时，双链分子的重组连接区要解旋，异源双链 DNA 的起始区可能不是常规的双螺旋形状，而是并排相联的两条链，这种类型的区域称为**平行连接**（praanemic joint）和一般双螺旋中典型的绞旋式不同。平行连接是不稳定的；进一步反应需要它们变成双螺旋型。这个反应与消除负超螺旋是同等的，可能需要一种酶通过产生一短暂的断裂，使链彼些盘绕来解决解螺旋和恢复螺旋的问题。

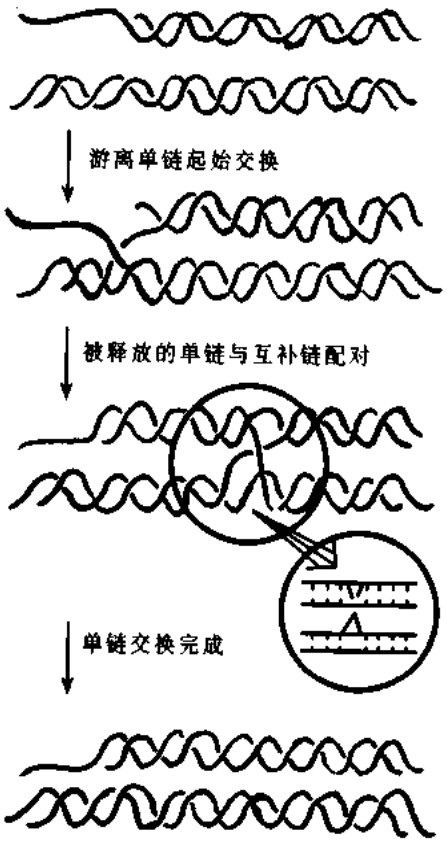


图 23－18　RecA 介导部分双链和完整双链之间的链的交换、产生的连接分子带有与重组中间体相同的结构。（仿 B.Lewin:Genes Ⅵ，1997，Fig.17－10）

到目前为止，所有的这些反应我们发现仅存在部分潜在的重组事件：一条单链侵入另一条双链。在 RecA 的作用下，两条双螺旋分子能彼此相互作用，它们有一个至少长 50nt 的单链区。这个单链区能使线形分子产生一个尾巴，或使环状分子产生一个缺口。

在部分双链和完整双链之间的反应导致了链的交换（图 23－18）。同化作用开始于线性分子的一端，在那儿侵入的单链以传统的方式取代了两链中与其同源的那条链。当反应到达两个分子的双链区时，侵入单链不和自己的互补链配对，而和另一条被取代链的互补链配对。

在这一阶段分子呈 *Chi* 型中间体结构。这个反应支持了在体外 RecA 能产生 Holliday 结构，表明酶能促使链的交互转移。我们不大知道由 RecA 结合而成的四链

中间体的几何学，但推测起来双链分子能以并排的方法来适合交换反应的需要。

在 *E.coli* 中有三组基因编码的蛋白和重组有关，*ruvA* 和 *ruvB* 的产物可促进异源双链结构的形成。RuvA 可识别 Holliday 连接的结构；RuvB 是一种 ATPase，它可发动迁移反应。图 23－19 表示复合体的功能。RuvA 结合在交换点上的所有 4 条链上。RovB 是六聚体，它围绕在交换点上游的每一条 DNA 双链的周围。

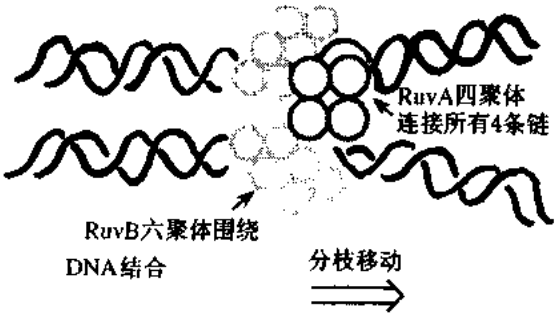


图 23－19　RUVAB 是对称的复合体，起动 Holliday 连接的分枝迁移。（仿 B.Lewin：Genes Ⅵ，1997，Fig.17－11）

此 RuvAB 复合体能导致分枝点以每秒 10～20bp 的速度迁移。RecG 解旋酶也具有相似的活性。RuvAB 解旋酶在反应时能将 RecA 从 DNA 中置换出来。RuvAB 和 RecG 都具有解离 Holliday 连接的活性，使重组得以完成。若二者发生突变，那么 *E.coli* 则完全失去了重组的活性。

第三个基因是 *ruvC*，它编码一种内切酶可特异识别 Holliday 分枝，它在体外可切这个连接体，解离重组中间体。一个共同的四核苷酸序列为 RuvC 提供了解离 Holliday 连接的热点，此序列是对称的，这样可以直接决定在哪一对链上产生缺刻，这样决定了补丁重组（并非全部重组）或剪接重组（在两个测翼标记向重组）的形成。

我们现在可以说明 *E.coli* 重组各阶段中不同蛋白所起的作用。图 23－20 表明在修复途径中细菌的酶催化重组的各阶段产生合适的 DNA 分子底物。但必须注意这个模型是细菌中 DNA 片段和完整染色体之间的相互作用，发生在受 DNA 损伤促进而进行的修复反应中，和真核生物中减数分裂期间染色体组的重组不完全相同。但 DNA 的操作和相同的分子活性有关。

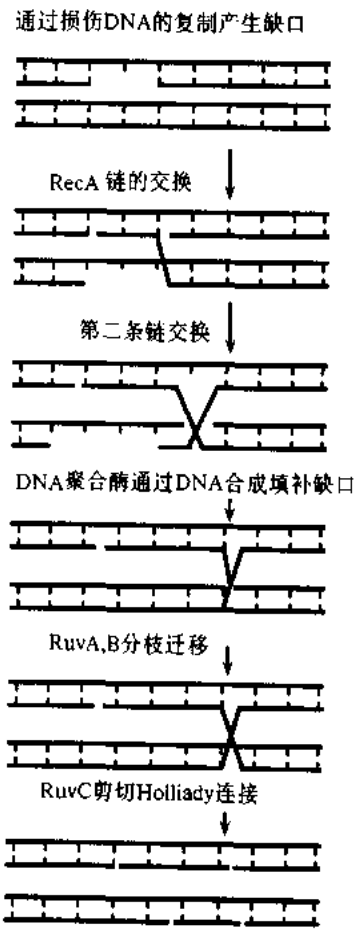


图 23－20　细菌的酶能催化在修复途径中的各重组阶段。（仿 B.Lewin：Genes Ⅵ，1997，Fig.17－12）

RecA 的同源蛋白在原核生物中普遍存

在,在真核中生物中也发现了相关的蛋白。在酵母中(*S.cerevisiae*) *DMC1* 和 *rad51* 两个基因所编码的蛋白和 RecA 相关。这些基因的突变会导致产生相似表型;突变体聚积双链断裂,且不能形成 SC。这些现象增强了一种观点:DNA 双链之间的交换和 SC 的形成有关,而且提出了染色体联会可能和细菌的同化反应相关,但在真核生物中 RecA 的同源蛋白并不形成"细丝",因此在真核生物中的反应机制可能和原核不同。

第五节　位点特异性重组

一、λ 在 *att* 位点的整合

位点特异性重组最典型的是 λ 噬菌体对 *E.coli* 的整合。λ在溶源周期和裂解周期中其 DNA 的物理状态不同。在裂解周期,λDNA 以环状分子独立存在于细菌的细胞质中;溶源化状态的 λDNA 整合在细菌的染色体中,称为原噬菌体。通过位点特异性重组这两种状态可以相互转变的。

进入溶源状态时游离的 λDNA 必须被**整合**(integrated)到宿主的 DNA 中;从溶源周期进入裂解周期时原噬菌体 DNA 须从宿主的染色体中切离出来。

在细菌和噬菌体的特异位点被称为**附着位点** (attachment sites, *att*) 上通过重组发生整合或切离。在细菌染色体上的特异位点称 att^1 ,若这个座位发生突变就会失去功能而抑制 λ 的整合。在溶源状态中原噬菌体人占据着 att^1。当 *E.coli* 染色体上 att^1 位点被消除时,一个感染的 λ 噬菌体可以通过整合到其他的地方来建立溶源化,但这个反应的效率只有在 att^1 位点整合效率的 0.1%。这种无效整合发生在第二个附着位点(secondary attachment sites),此位点和 att^1 位点很相似。

细菌的 att^1 位点称有特异序列为 *att* B,其序列的成分是 BOB′,在噬菌体 *att* 位点上的特异序列,称为 *att P*,含有 POP′。图 23-21 表示在这些位点上发生的重组。"O"序列是 *att B* 和 *att P* 共有的,被称为**核心**(core)序列,长 15bp,重组就发生在此序列上。两侧序列 B,B′和 P,P′是作为**臂**(arms);臂序列各不相同。由于 λDNA 是环状的。重组时它插入到细菌染色体中呈线状。原噬菌体两端到 att 位点是由重组产生的 BOP′和 POB′。

构建 *att* 位点的一个重要结论是整合和切离并不涉及反应序列的相同部分。整合需识别 *att P* 和 *att B*;而切离时需要 *att L* 和 *att R*。位点特异重组是通过重组位点的鉴别来控制的。

虽然这种重组事件是可逆的,但不同的条件决定了反应的方向。这在噬菌体生活周期中是一个重要的特点。噬菌体一旦整合进入溶源周期不会立即被切离,反之也然。

整合和切离反应作用位点是不同的,因此这两个反应所依赖的蛋白也不相同。

整合 (*att B* × *att P*) 反应需要噬菌体 *int* 基因的产物和细菌蛋白宿主整合因子(integration host factor,IHF)。

切离(*att L* × *att R*)反应需要噬菌体 Xis 基因的产物以及 Int 和 IHF。

Int 和 IHF 是两个反应都需要的,Xis 在方向控制上起到重要的作用。它是切离所需要的,对整合起抑制作用。

IHF 是 20kDa 的蛋白,由两个不同的亚基组成,分别由基因 *himA* 和 *himD* 编码。*IHF* 并不是 *E.coli* 中主要蛋白,它并不参与细菌的同源重组。它是包被在 DNA 外表上

的几种蛋白之一。*Him* 基因的突变就会阻断 λ 的位点特异性重组。重组也能被 *int* 突变所抑制，表明 IHF 和 Int 可相互作用。

位点特异重组能通过 Int 和 IHF 在体外来完成。它涉及精确的断裂和重接而不需要任何 DNA 合成。*att* 位点的作用通过制备其两侧中的一侧缺失来进行研究。结果表明 *att P* 远长于 *att B*。*att P* 的功能需要一个 240bp 序列，而 *att B* 仅需 23bp 的片段（-11 ~ +11）就具有功能，两者之间有 15bp 是同源的，称为核心区。在核心区中有 7bp 是交错剪切的序列。*att B* 和 *att P* 的大小就表明它们在重组中所起的作用是不同的。*AttP* 提供了附加的信息，以和 *att B* 相区别。

att P 和 *att B* 序列是通过什么样的相应机制来进行切割和交换呢？是同时切割交换（协同剪切），还是经单链交换先产生 Holliday 连结，再进行环的裂缺和连接进行重组（连续交换）呢（图 23 - 22）？通过用“自杀性底物”将重组反应停止在中间体状态，在此结构中核心序列被切开。缺刻的存在干扰了重组的进行。这样可使重组开始进行但又不能完成，从而便于鉴别。这些中间分子的结构表明通常是发生单链交换。Int 蛋白解离 Hollidy 连接，它负责剪切和连接。

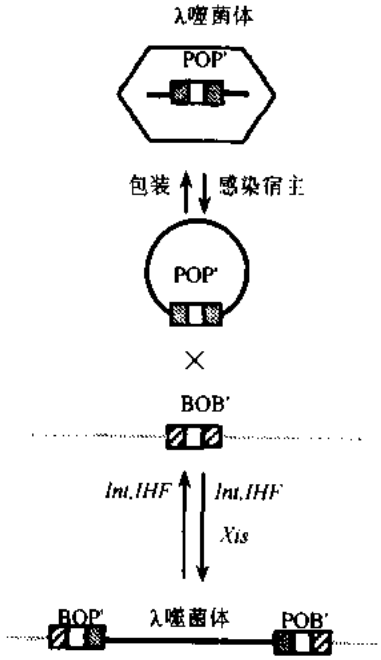


图 23 - 21 λ 噬菌体的整合和切离。（参考 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 19 - 22）

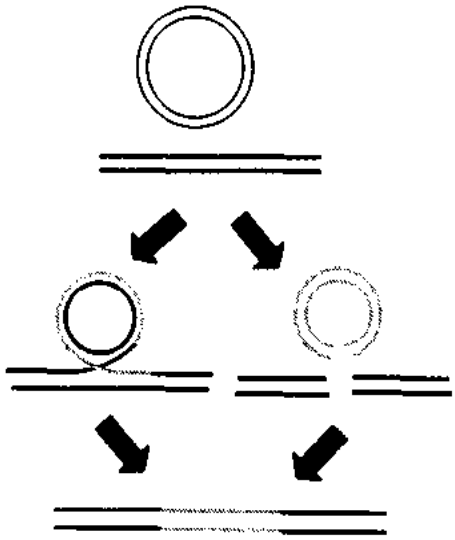


图 23 - 22 *attP* 和 *attB* 之间的重组是连续交换，还是一致性剪切？（仿：B. Lewin: Genes Ⅵ, 1997, Fig. 17 - 19）

图 23 - 22 表明位点特异重组模型。若 *att P* 和 *att B* 位点经相同的交错切割，单链末端能交叉杂交。在距 λ 交换位点 7bp 处进行反应并产生 3′ - P 和 5′ - OH。这个反应是 λ 和 *E. coli* 产生相同的交错单链末端，交互退火，其实质是同族重组。各自双链上的对应链在相同的位点被剪切，在双链间彼此交换 3′游离端，分枝点在 7bp 的同源区中移动，然后这个结构在另一对对应链被解离。

在体外这个反应要求在 *att P* 中超螺旋的存在，但在 *att B* 中却没有超螺旋。在体外两个超螺旋 DNA 分子之间进行反应时，几乎所有的超螺旋都保存在整合产物中，因此反应过程中不可能存在单链旋转的游离中间体。因此认为反应的过程是通过一个

Hollidy 连结进行。这个断裂和愈合反应类似于拓扑异构酶 I 的活性，所不同的是来自不同双链的断裂端彼此连接，而不是同一条双链的断裂端的重接。Int 具有 TopI 松弛负超螺旋 DNA 的能力。

在体外重组中需要大量的 Int 和 IHF 蛋白。Int 和 IHF 协同地结合于 *att P*，它们的对于此位点亲和性可通过超螺旋来增强。高化学剂量表明这些蛋白并不具有催化功能，而是提供某种结构维持单个重组。

这两种蛋白在位点特异重组中涉及 *att* 区域特异位点的结合。Int 有两种不同的结合模型。它在核心序列反向结合切割位点，在每条单链上切割(图 23－23)。这些位点共享一致序列。Int 也结合 *att P* 臂上的位点，其一致序列不同。Int 不同的功能区识别

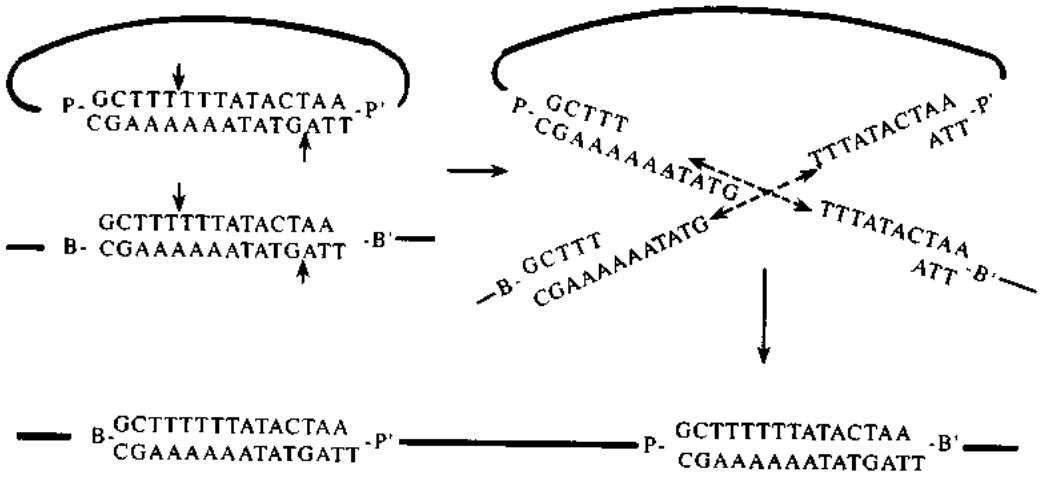


图 23－23　在 *attP* 和 *attB* 的共同核心序列中的交错切割使"＋"字型的配对产生交互重组连接。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 17－20)

各类序列：N 端功能区识别 *att P* 的臂，C 端功能区识别 *att P* 和 *att B* 的核心区。这两个功能区能同时结合 DNA，这样就可以将 *att P* 的臂拉向核心区。

IHF 结合 *att P* 中约 20bp 的序列，结合位点和 Int 的结合位点邻近。Xis(excisioase)有两个结合位点，在 *att P* 中互相靠近，保护区域超过 30～40bp。Int，Xis 和 IHF 蛋白几乎覆盖了 *att P* 的所有区域。Xis 的结合改变了 DNA 的结构，使其成为整合反应的惰性底物。而促使反应向切离方向进行。

图 23－24 表明核心序列被 Int 结合时和 DNA 的各连接点都在一个表面上，两个切点位于大沟。IHF 的结合位点位于同一表面上；若 Int 结合位点和例翼的 IHF 结合位点之间的间隔改变，那么整合的双螺旋数就不同，整合将被抑制。

当 Int 和 IHF 结合于 *att P* 时，在结合位点上产生一种复合体。它们在蛋白寡聚体表面上互相牵拉，*att P* 的超螺旋对于形成**整合体**(integrant)是必要的。

在 *att B* 上仅有的一个结合位点是在核心区的两个 Int 位点，但对游离 DNA Int 并不能直接结合在 *att B* 位点。此结合体是一个捕捉 *att B* 的中间体。Int 分子可能结合在 *att B* 核心区的结合体的一部分 (图 23－25)。

根据这个模型在 *att P* 和 *att B* 之间的起始识别并不直接依赖于 DNA 同源性，而取决于 Int 蛋白识别 *att* 序列的能力。然后这两个 *att* 位点通过整合体预先决定方向被拉到一起。在这一阶段同源序列是重要的，因这时链交换反应是必须的。

若加上 Xis 的话，Int 和 *att R* 也能形成相似的复合体，这个复合体能和 Int 和 *att L*

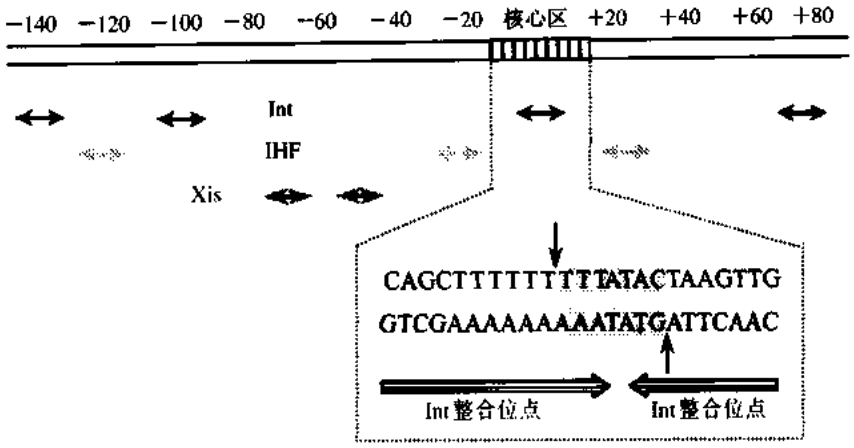


图 23－24　Int，IHF 和 Xis 在 attP 上的结合位点。（仿 B.Lewin：Genes Ⅵ，1997，Fig.17－21）

形成的一个凝缩复合体配对。这个反应是不需要 IHF 的，以上事实表明了整合和切离反应的对称性。

很多位点特异重组复合体都是调节反应所需的，所以当病毒进入溶源状态时整合先进行；当原噬菌体进入裂解周期时，先进行切离。通过 Int 和 Xis 量的控制将进行相应的反应。

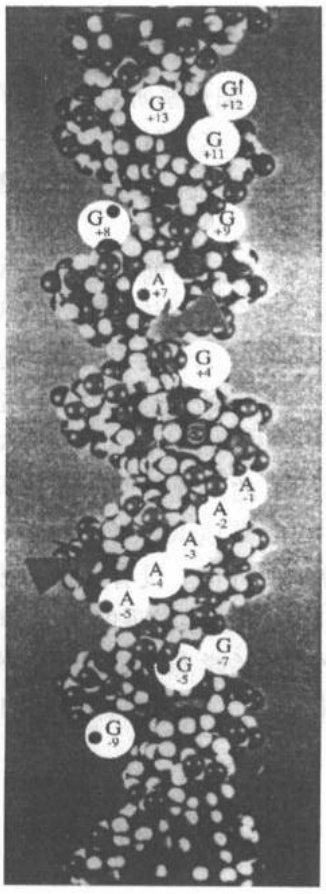


图 23－25　Int 在核心区的结合位点是在 DNA 的同一表面上。（仿 B.Lewin：Genes Ⅵ，1997，Fig.17－22）

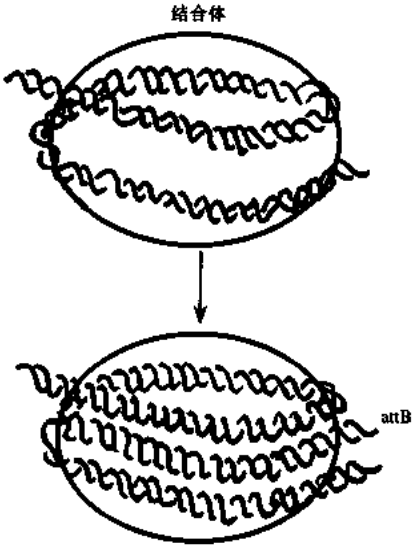


图 23－26　多拷贝 Int 蛋白可将 attP 组入结合体中，通过对游离 DNA 上 attB 的识别起始位点特异性重组。（仿 B.Lewin：Genes Ⅵ，1997，Fig.17－23）

二、倒位重组

沙门氏细菌（*Salmonella*）的相转变（图 23－27）和 Mu 噬菌体 G 片段的移位（见第 16 章）都属于位点特异性重组。沙门氏菌的 *hin* 基因长 995bp，它和 Mu 噬菌体的 G 片段的相应基因 *gin* 同源性很强，功能也能在互补实验中互相替代。P_1 噬菌体中决定 C 片段倒位的 *cin* 基因也和 *hin*，*gin* 同源。转座子 Tn*3* 的 $tnpR$ 基因产物和 Hin 也有一定的同源性。Mu 噬菌体 G 片段两端 34bp 的反向重复序列和 Hin 片段的 IRL 和 IRR 也是高度同源的。倒位的机制是 IRL 和 IRR 互相配对，形成一个环，Hin 蛋白介导它们重组。切离和重组方式和 Tn_3 共合体解离相似（见本章）。

三、交配型转变的重组

酵母交配型的转变这种重组有以下的特点：①需要特殊蛋白 H0；② 需长同源序列；③位点是特异的，仅在 Y 区重组；④ 重组时产生 *Chi* 结构；⑤H0 识别特殊位点并交错切割；⑥复制过程中涉及修复性复制。以上特点表明：这种重组既不同于同源重组，也不同于位点特异性重组，因此应单列为另一类型，称为**同源特异重组**。

第六节　原核生物的转座因子

1951 年美国遗传学家 McClintock 根据对玉米染色体的长期观察研究，提出了**转座**（transposition）的概念，也就是说基因可以在染色体上移动，这当时几乎没有一位遗传学家能接受这种超越时代的新概念，直到 1983 年，这位 80 多岁的老妇人终于登上了诺贝尔奖的领奖台，到此时她的重大发现几乎被埋没了 30 年。

转座是遗传学发展史中最为重大的发现之一，把基因概念向前推进了一大步。McClintock 的伟大就在于她以传统的工具和方法，却获得了分子遗传学中划时代的突破；没有丰富的想象力，没有创新精神是难以达到的。Mendel 的自由组合和独立分配二大定律的发现，Morgan 的连锁定律的建立，Watson 和 Crick 的双螺旋模型都是划时代的发现，但他们都是在前人工作的基础上获得的，除 Mendel 外，Morgan 和 Watson 的发现都是由多位学者合作而得以攻克的。而 McClintock 的发现可谓前无古人，无可借鉴，同时又是她一人单枪匹马，孤军作战。她不仅默默地承受工作的压力，还要承受舆论的压力，直至得到广泛的承认，这也十分难能可贵。没有坚韧不拔的精神是难以坚持的。现在人们用分子生物学的方法已经确证她早年发现的 *DsAc* 等系统都是客观存在的，完全正确。的没有求实的精神，严谨的治学态度难以达到的如此准确的预见。McClintock 这种大胆设想，小心求证的精神是值得任何自然科学工作者学习的楷模。

一、原核生物中的转座因子的发现和检出

虽然 McClintock 早在 1951 年就在玉米中发现了转座因子，但由于人们受到基因在染色体上有序排列的传统观念的桎梏而难以接受“**跳跃基因**”这一新的概念，所以直到 1967 年 Shapiro 才在 *E. coli* 中发现了**转座因子**（transposable element）。

他在半乳糖操纵子(见 17 章)中发现了一种**极性突变**。*gal* 操纵子有三个基因:*gal K, T, E*,它们编码三种酶(图 23－28)催化 Gal 的分解代谢。其中间产物 Gal－1－p 是有毒的,因此 *galT*- 在含半乳糖培养基上因使 Gal－1－p 积累而难以成活。那么在此培养基上能生长的各种突变型,应是 $galT^-$ 的回复突变及 $galK^-\ galT^-$ (因它不能产生 Gal-1-p),但意外发现 $galE^-\ galT^-$ 也能生存,按理说这种突变型可使 Gal-1-p 积累是不可能存活的,但居然生长良好,其原因何在呢? 经进一步研究发现 $galE^-$ 是极性突变,由于它的突变使远离操纵位点的 *galK* 活性大大下降,因而细胞中不会积累 Gal-1-p 之故。

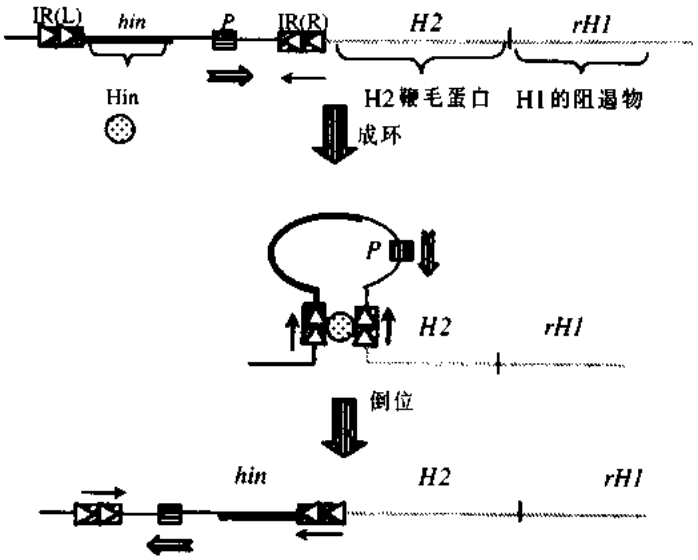


图 23－27　沙门氏菌相转变的位点特异性重组。

以上的这种极性突变又和一般的极性突变不同,它有以下的特点:① 能回复突变,表明不可能属于缺失或移码突变;②用诱变剂处理并不能提高回复突变率,因此推测可能不是点突变。既排除了点突变,缺失和移码突变,那么还有什么样的变异能引起极性突变呢? 此时人们已知道了 F 因子和 λ 噬菌体的整合功能,因此他们想到 $galE^-$ 的突变可能也是部分 DNA 的插入(突变)和切离(回复突变)所致。接着他们设计了一系列实验来证实这一想法。

(1)密度梯度离心实验

E. Jordon 等和 P. Starlinger 小组以及 J. Shapiro 分别进行了以下实验,他们将含有 gal^+ 和 gel^m (上述极性突变) λ 噬菌体分别分离出,同时加入到离心管中进行 CsCl 密度梯度离心,结果出现了两条带,$\lambda dgal^m$ 的密度 > $\lambda dgal^+$,表明 $\lambda dgal^m$ 有可能具有小片断 DNA 的插入。

(2) 分子杂交实验

将 $\lambda dgal^m$ 和 $\lambda dgal^+$ 进行分子杂交,若 $\lambda dgal^m$ 有额外片段那么在电镜下可以观察到它们存在和所处的位置。实验结果在杂交链上出现了一个额外的单链的茎环结构,长 800nt(图 23－28)这就是**插入序列 1**,或称为 **IS1**(insertion sequence 1)。

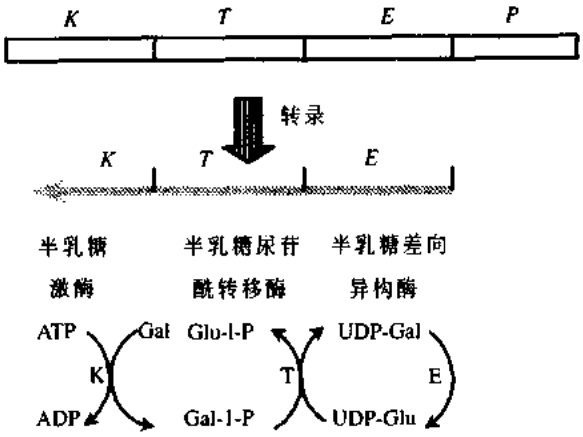


图 23－28　*gal* 操纵子及 K,T,E 催化的 Gal 分解代谢。

二、原核转座因子的类型

1. 插入序列(IS)

IS*1* 是基因组中可移动的遗传因子家族中的成员之一，它可以整合到宿主非同源位点上，这就是**转座**(transposition)，若 IS 插入到某基因内，通常这个基因就会失活。这种精确的作用取决于 IS 有关的状况。例如 IS2 因子以同一方向插入到染色体中，则会减少基因的表达，以相反方向整合，则会增加基因表达。而 IS*1* 因子不论以什么方向插入都会降低基因的表达。

IS 原来分为 IS 1～4，后面的数字反应它们被分离的先后次序，和转座因子的总数无关。

正常的细菌染色体和质粒都含有 IS。*E.coli* 的标准品系似乎含有各种拷贝数(＜10)的 IS 因子。在原核生物中 IS 家族有很多成员，它们的结构相似，两端都有短的正向重复序列(direct repeats，DR)，略长的**反向重复序列**(inverted repeats，IR)以及 1kb 左右的编码区，它仅编码和转座有关的**转座酶**(图 23－29)。对靶的选择有三种形式：**随机选择，热点选择和特异位点的选择**。

由于 IS 中有反向重复序列的存在，因此如果质粒上有 IS，那么经变性和复性后，在电镜下可以观察到茎环结构的存在(图 23－30)。

IS 的转座是由转座酶(transposase)催化的，它由 IS 编码。首先转座酶交错切开宿主

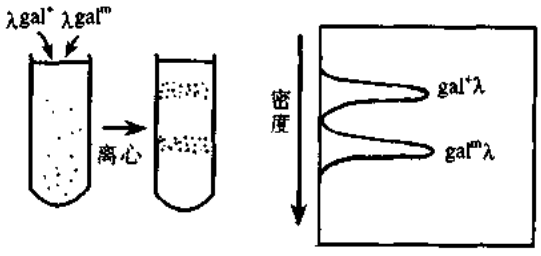


图 23－29a　λ*dgal*^{m} 的密度梯度离心实验。

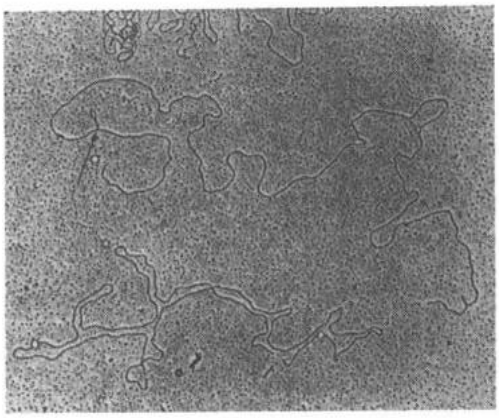
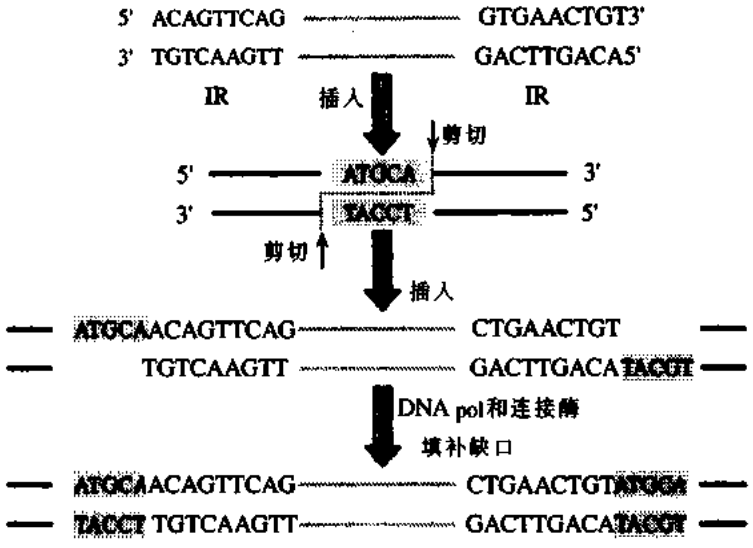


图 23－29b λdgal⁺ λdgal⁻ DNA 异源双链的电源照片。箭头指向的单链环是 λdgal⁻ 中存在插入顺序所致。
（引自 Griffiths A. J. F. *et al*.: An Intiroduction to Genetic Analysis, 5th ed., 1993, Fig. 13－8）

靶位点，然后 IS 插入，与宿主的单链末端相连接，余下的缺口由 DNA 聚合酶和连接酶加以填补，结果使插入的 IS 两端形成了 DR 或靶重复（图 23－30）。



	靶顺向重复 (bp)	反向重复 (bp)	转座子长度(bp)	反向重复靶的选择	靶顺向重复拷贝数
IS1	9	23	768	随机	5~8
IS2	5	41	1327	热点	5(在F中为1)
IS4	11~13	18	1428	$AAAN_{20}TTT$	1或2
IS5	4	16	1195	热点	,?
IS10R	9	22	1329	NGCTNAGCN	
IS50R	9	9	1531	热点	
IS903	9	18	1057	随机	

图 23－30 IS 的结构和性质。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18－1）

2. 类插入序列(IS-like elements)

是指 IS*10R*,IS*50*R 和 IS*903*,它们的结构和 IS 相似(表 23－4),但不独立存在。而是作为**复合转座子**(composite transposons)两臂的组件。

表 23－4　IS 的结构和功能

	DR(bp)	IR(bp)	中心区域(bp)	靶的选择	拷贝数
IS 1	9	23	768	随机	5～8
IS2	5	41	1327	热点	5 (在 F 因子上为一)
IS4	11～13	18	1428	AAN20TTT	1 或 2
IS5	4	16	1195	热点	?
IS10R	9	22	1329	NGCTNAGCN	
IS50R	9	9	1531	热点	
IS903	9	18	1057	随机	

参考 B. Lewin: Genes Ⅵ, 1997, Fig18.1

3. 复合转座子

复合转座子要比 IS 长得多,中心区域编码抗性标记。不同的复合转座子的抗性标记不同。复合转座子两端的组件由 IS 和类 IS 组成。有的二侧组件相同(如Tn*903*),有的不同(如 Tn*10*)。有的方向相同(如 Tn*9*),有的方向相反(如 Tn*903*, *10*, *5*)。有的两侧组件皆有功能(如 Tn*903*, *10*),有的仅右侧组件有功能(图 23－31)。

据推测当两个独立的组件与中心区域连接时发展成复合转座子。这个结构可能由 IS 因子转座到紧靠供体位点的受体位点上。这两个组件可能相同或存在差异。单个组件转座整个复合转座子的能力表明两个组件保持活性不受选择压力的影响(表 23－5)。

表 23－5　复合转座子的结构和功能

转座因子	长度(bp)	遗传标记	末端组件	方向	二组件的关系	组件的功能
Tn*903*	3100	Kan^R	IS903	反向	相同	二者皆有功能
Tn*9*	2500	Cam^R	IS 1	正向	推测相同	预计有功能
Tn*10*	9300	Tet^R	IS 10R			有功能
			IS 10L	反向	有 2.5%的差异	无功能
Tn*5*	5700	Kan^R	IS 50R			有功能
			IS 50L	反向	1 bp 的改变	无功能

B. Lewin: Genes Ⅵ, 1997, Table 34.2

在复合转座中是什么替代单个组件而担负起转座的作用呢? 特别是在两个组件都有功能的情况下。在 Tn*9* 例子中,它的组件是 IS*1* 因子,推测它们都有功能。为什么转座是依赖整个复合转座子,而不是 IS 的本身呢?

两个 IS 因子能使夹在其中的任何序列转座。如一个环状复制子上的 Tn*10* 它的两个组件 IS*10*L 和 IS*10*R, 可以看成是复制子序列两端的组件, 因此在进一步转座时

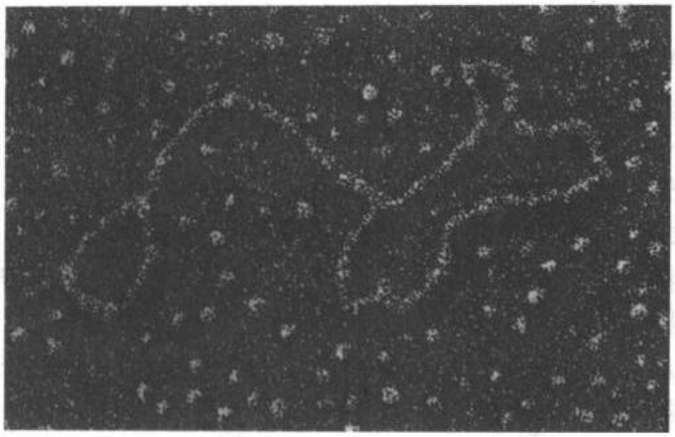


图 23－31　当质粒 DNA 变性时出现了特殊的结构。大环和小环之间的区域是双链 IR。
（引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 20－7）

它不仅可以使 Tn*10* 转座，也可使复制子序列转座，但两侧组件的序列与在 Tn*10* 中不同。（图 23－32）。

4．TnA 家族

转座子（transposon，Tn）另一家族是 TnA，长约 5kb 左右，两端具有 IR，而不是 IS，中部的编码区不仅编码抗性标记，还编码转座酶和解离酶。

TnA 是复制转座的转座子。在此家族中 Tn*3* 和 Tn*1000*（γ δ）研究得最深入。通常末端有 38bp 左右的 IR，在两个 IR 中任一个顺式作用缺失都会阻止转座。5bp 的正向重复是靶位点产生的。它们都带有抗性标记（图 23－33）。

TnpA 介导转座分为两步，分别由 *tnpA*、*tnpR* 编码的转座酶（transposase）和解离酶（resolvase）来完成的。这两个酶是通过突变鉴别出来的。此类转座和末端的 IR 有关。解离序列（res）是内部的特殊位点，只有 TnA 家族才具有这一位点。

tnpA 突变体是不能转座的，这个基因的产物是一种转座酶。它可以结合在末端 38bp

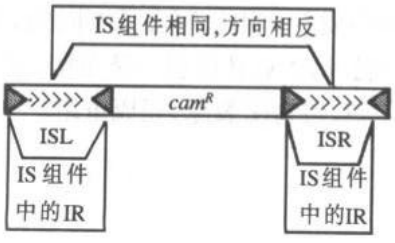


图 23－32　复合转座子结构（Tn9）。

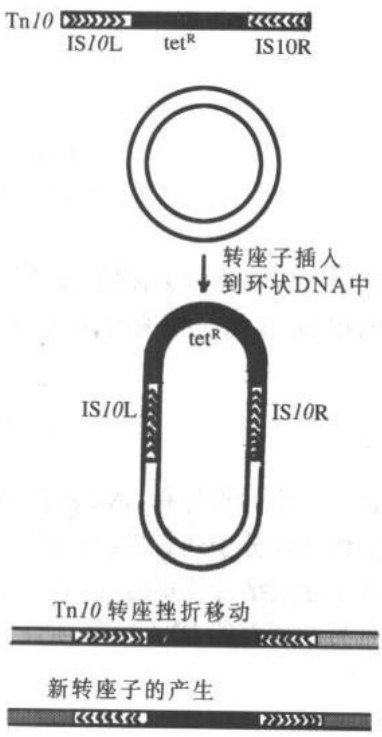


图 23－33　两个 Tn*10* 组件也能使它们之间的其他顺序转座。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 17－21）

IR 中的 25bp 的序列上。*E. coli* IHF 的结合位点就在转座酶结合位点附近(在靶位点附近)。而转座酶结合和 IHF 的结合是协同性的。转座酶识别末端重复序列,并可以使靶 DNA 产生 5bp 的交错切割使转座子插入。IHF 是一种 DNA 结合蛋白,常在 *E. coli* 中组装成一个大分子,它在转座反应中并不起作用。

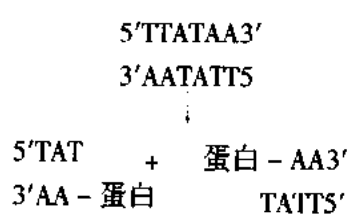
tnpR 的产物具有双重功能,一是作为基因表达的阻遏物,另一个作用是解离酶的功能。*tnpR* 突变将增加转座频率,这是由于 TnpR 阻遏了 *tnpA* 和它自己的基因转录,TnpR 蛋白失活就使 TnpA 合成增多,结果增加了转座的频率。这表明 TnpA 转录酶的数量是转座的限制因子。

tnpA 和 *tnpR* 基因之间有一个富含 A－T 的**内部顺式控制区**(图 23－33),TnpR 的两种功能就是通过和此区的结合来实现的。

TnpR 作为解离酶介导 Tn*3* 在共整合结构中正向重复之间的重组。一个**共合体**(cointegrate)结构原则上通过同源重组被解离开。操纵子突变作图正好在此位点的左侧。这个位点如何相互作用呢?一种可能就是 TnpR 结合在这三个位点上可以使 DNA 产生适当的拓扑结构,而结合在这一套位点上,就可能会阻遏 *tnpA* 和 *tnpR* 的转录。

在体外解离是用一个类共合体结构的分子作为底物来进行分析。这个底物必须是超螺旋的。它的解离产生了两个串联的环,每个含有一个 *res* 位点。这个反应需要大量的 TnpR 解离酶,而不需要宿主的因子。解离发生在大量核蛋白结构中。解离酶结合在各个 *res* 位点上,然后将这些位点聚集在一起形成一个约 10nm 的二聚体结构。在反应时超螺旋发生改变,在 *res* 位点 DNA 被结合的转座酶弄弯了。

解离是一种非复制性反应,键的断裂和重接不要求能量的输入。在共合体解离中产生相同的中间体阶段,是在 res 位点切开后产生的两条双链,其 5′端与解离酶共价结合。对称剪切发生在短的回文序列上,产生了 2 个碱基的单链突出。把这个观点扩展到 I 位点的交换区,我们就能发现剪切反应是:



这个反应和 λ 噬菌体的 Int 在 *att* 位点的作用相似,而 20bp 的 *res* 位点中的 15 碱基序列和 *att* 位点是相关的,我们把这两个剪切位点并列排列如下:

res GATAATTTAT↓AATAT
att GCTTT↓TTTATACTAA

这个反应的本身和 DNA 操作类似,然而解离仅发生分子内的位点,而 *att* 位点之间的重组是在分子间直接进行的。这两种反应中蛋白质的作用机制也是不同的。解离酶以 4 个亚基结合在重组 *res* 位点来发挥自己的功能,每个亚基剪切一条单链,然后亚基识别另一相对的 DNA 单链,把它们放到重组构象中,再封闭裂隙,完成重组。

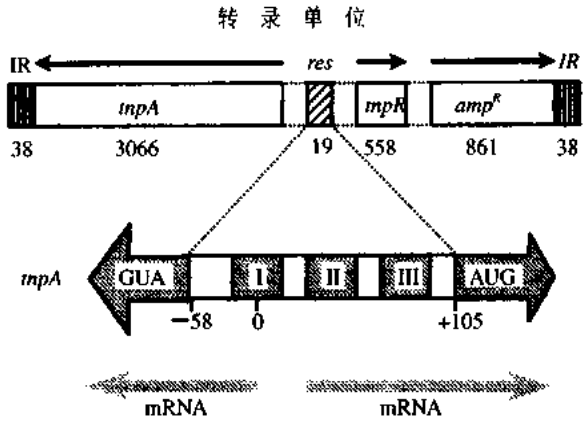


图 23－34　TnA 的结构和转录。(参考 B.Lewin: Genes Ⅵ, 1997, Fig. 18－16)

三、转座的类型

转座子插入到 DNA 上新的位点，首先交错切开靶 DNA，再将转座子连接到靶的凸出单链上，最后填补空缺完成转座。在两个单链上交错切割决定了正向重复的长度，因此各种转座子重复序列的描述反应了在剪切靶 DNA 有关酶的拓扑学。

参差不齐的末端的使用是各种转座子的共同特点，我们可以将转座分成三种不同的类型：

复制型转座（replicative transposition）　这种类型的转座因子在反应中产生重复。因此转座的整个 DNA 序列是原转座因子的一个拷贝，而不是它本身。图 23－35a 表示这种转座的结果。转座子作为移动的拷贝，一个拷贝留在原处，另一拷贝插入到新的靶位点。转座伴随着新拷贝的复制。这种类型的转座涉及到两种酶：一是转座酶，它要作用原来转座子的末端；二是解离酶，它作用在重复的拷贝上。与 Tn*A* 相关的一群转座子仅通过复制转座进行移动。

非复制型转座（nonreplicative transposition）　这类转座因子作为一个物理的整体直接从一个位点移到另一位点。这种类型又有两种机制：一种是利用供体和受体靶 DNA 序列的连接，有的步骤和复制型转座相同。插入序列和复合转座子 Tn*10* 及 Tn*5*

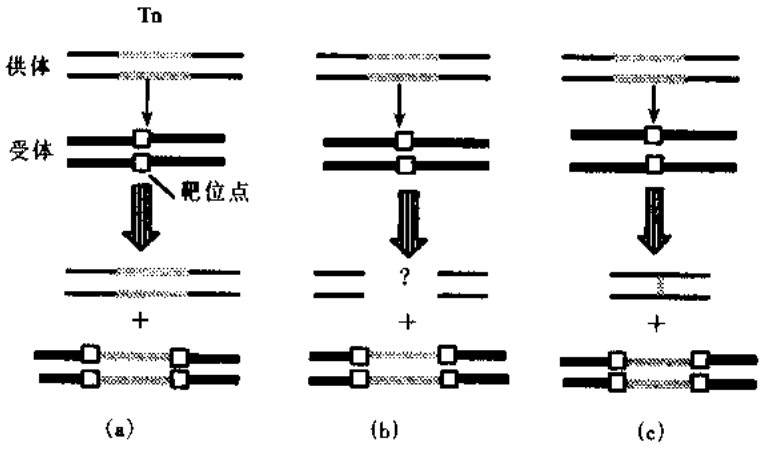


图 23－35　三种不同的转座机制：(a) 复制性转座；(b)非复制性转座；(c)保守型转座。

就是用这种机制转座(图 23 - 34b)。这涉及到一个问题就是转移时转座子怎样从供体 DNA 上释放出来。这种类型的机制只需要转座酶。另一类非复制型转座是**保守转座**。两种非复制转座的机制使转座因子插入到靶位点，而供体原来的位点已不存转座因子。非复制转座之后供体分子发生了什么情况呢?一种可能是供体消失了(可能细菌可耐受多个拷贝);另一种可能是宿主的修复系统能识别此处的双链断裂而将它修复。

保守转座(conservative tranposition)　保守转座是属另一类非复制型转座。这种类型的转座因子从供体位点上切离，插入到靶位点上，在一系列的反应中供体上转座子两侧的 DNA 双链都被保留着(图 23 - 35c)，此和 λ 噬菌体的整合机制十分相似。而且这种因子的转座酶和 λ 整合酶家族相关。采用这种机制的转座因子较大，不仅它本身可以转座，且可介导供体 DNA 从一些细菌转移到另一些细菌中。虽以前把它们归类于转座子，看来作为**附加体**(episome)更为适合。

某些转座子只具有一种转座机制，而另一些可能具备两种途经。IS1 和 IS903 就采用复制和非复制途径。Mu 噬菌体能从共同的中间体转变为两种途经中的一种。

另外简单的分子内转座是新位点的插入，这是转座子启动另一类型的 DNA 重排，这类事件中有的是转座子多个拷贝之间亲缘关系的结果。另一些存在着不同的转座机制交替使用，难以划分它们的类型。

宿主 DNA 的重排可能是一个转座子将一个拷贝插入到原来位置附近的第二个位点内。宿主系统可能负责转座子两个拷贝之间的交互重组;结果取决于重复序列的存在以及它们的排列方向。

图 23 - 36 表示在两个 DR 之间的重组使夹在两个 DR 中间的 DNA 序列被切离而缺失，间插区形成了一个环(将从细胞中丢失)，DR 在染色体上只留下一个拷贝。应注意复合转座子 Tn9 的两端组件 IS1 是正向排列的(表 23 - 3)，它们之间的重组将使 Tn9 原来的位点上留下了单个 IS1 组件的转座子。

转座子附近序列的缺失是两步反应产生的；转座产生了转座子的正向重复，而重组又发生在此 DR 之间的转座子分子内。当新拷贝的方向和原拷贝方向相反时，重组会导致倒位；当方向相同时重组的结果产生了两个小的环状 DNA，其中一个无复制原点而被丢失，另一个有复制原点而可以复制。因丢失的环带有转座子附近的序列，结果造成转座子附近序列及一个重复序列的缺失(图 23 - 37)。

另一个反向重复序列之间交互重组的结果是夹在两个 IR 之间的 DNA 发生倒位(图

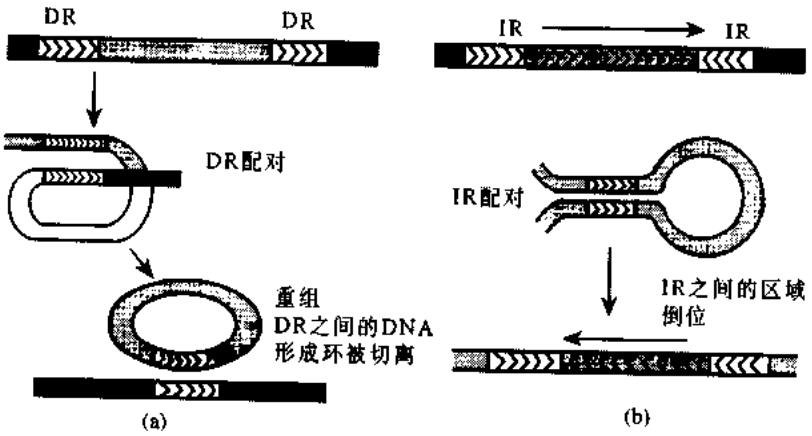


图 23 - 36　DR 之间和 IR 之间的重组结果。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18 - 8, 18 - 7)

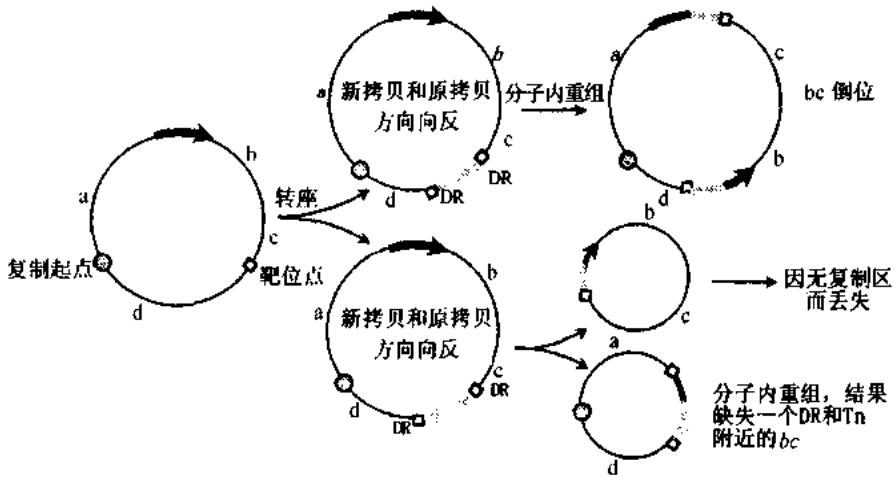


图 23－37　转座子分子内重组产生的倒位和缺失。（参考 D. Freifeldel；Molecular Biology，1983，Fig. 19－20）

23－35)，而反向重复可以得到保留，这有利于引起下次的倒位。应注意复合转座子即使与组件相关的中心区域通过重组可能被倒位，但它仍是基因组中稳定的组分。

有的转座子可以促进**双重倒位**(duplicative inversions)。这些转座子以相反的方向位于中心区的任一侧，中心区发生倒位时，它们的方向也随之改变。

转座的本身并不需先通过切离，可能是细菌的酶识别转座子中的同源区域。这是很重要的，因为缺乏这种酶转座子位点插入的功能就消失了。**精确切离**(precise excision)是使转座子从原来位置上消失，其结果是使发生插入突变的基因发生回复突变。这是很罕见的，Tn*10* 发生的频率为 10^{-6}。在 IS*10* 组件中切离涉及到 24bp 序列之间的重组。这些序列是反向重复，但由于 IS*10* 组件本身被倒位，而在 Tn*10* 中形成了正向重复(图 23－38)。

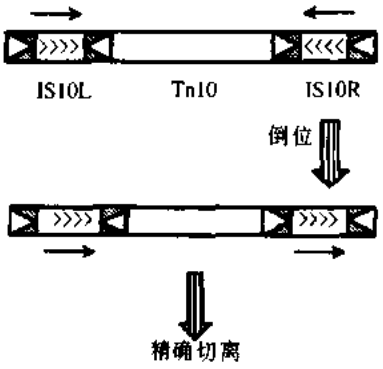


图 23－38　Tn*10* 的精确切离。

这一个大序列的不精确的切离和精确切离相比较反应了 DR 长度的增加(精确切离的正向重复为 9bp，非精确切离为 24bp 长)。无论哪一种切离都依赖转座子的编码功能。但机制不清。切离是 *RecA* 非依赖性的，而且使人们联想到细胞中间隔很近的重复序列之间的自发缺失。

四、转座的共同中间体

很多的可移动 DNA 因子都是通过基本相似的机制从染色体上的一个位置移到另一位置。它们包括 IS，原核和真核转座子以及细菌噬菌体 Mu。反转录 RNA 的 DNA 拷贝的插入也是用同样的机制。免疫球蛋白重组的第一步与之相似。

转座开始以共同的机制使转座子和它的靶顺序相连接。图 23－39 表示转座子的两端和靶序列在两条单链上被交错切割，二者切口末端交叉共价连接。转座子的两端先进行剪切，随之末端联会，使转座子的两个末端结合在一起。现在搞得最清楚的是 Mu 噬菌体，它以两种方式利用此转座过程。它感染宿主细胞后通过"非复制转座"插入

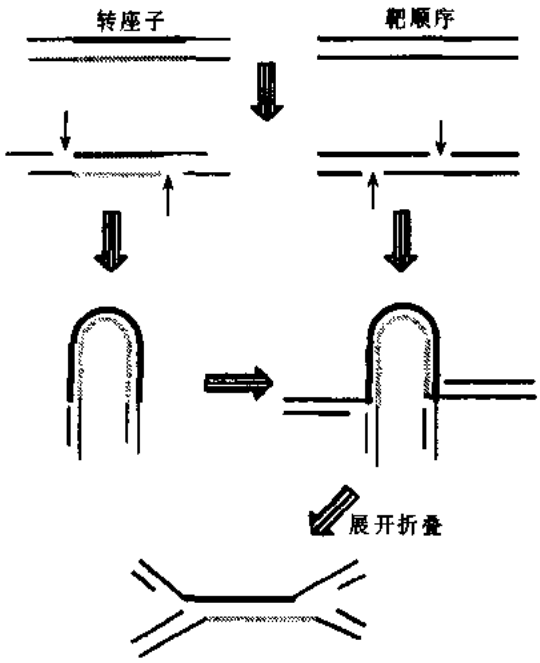


图 23－39　切口起始转座。
（仿 B.Lewin: Genes Ⅵ, 1997, Fig. 18－10）

到宿主基因组中。在进入裂解周期时，通过复制型转座扩增大量的拷贝。两种类型的转座都涉及转座子和靶序列之间的相同的反应，但随后的反应不同。

Mu 噬菌体 DNA 的起始操作是由 MuA 转座酶完成的。带有 22bp 一致序列的 3 个 MuA 结合位点位于 DNA 的两端。L1、L2 和 L3 在左端；R1，R2 和 R3 在右端。MuA 的单体可各结合一个位点。MuA 也能和噬菌体基因组内部位点相结合。这样它在左右端及中部形成一个复合体。内部位点的作用并不清楚。但它的存在对于形成复合体是需要的，但对单链的剪切和以后的反应不起作用。图 23－40 表示 Mu 转座子 DNA 和靶序列通过 3 个阶段反应进行连接。这只涉及转座子两端的 2 个位点。MuA 亚基结合在这些位点形成一个四聚体，这样产生了转座子两端的联会。这个四聚体现在的功能是使 MuA 的两个末端同时进行反应。MuA 有 2 个 DNA 操作位点，它们的作用方式是强迫转座酶的亚基进行反式作用。一致序列的结合位点与 22bp

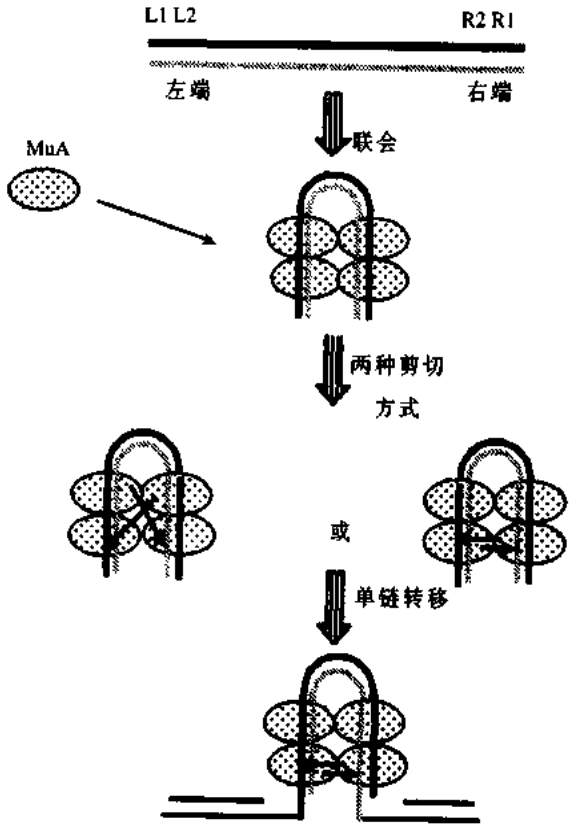


图 23－40　Mu 转座 DNA 和靶顺序的连接。（仿 B.Lewin: Genes Ⅵ, 1997, Fig. 18－11）

的一致序列 L1,L2,R1 和 R2 相结合。这个活性位点在 Mu DNA 单链上剪切,剪切位点是在 MuA－结合位点 L1 和 R1 附近。但它也能剪切 DNA 不同链上的某些序列。

此转座子的末端被反式作用的 MuA 亚基剪切。这种反式作用是指结合于 L1 和 R1 的 MuA 单体并不剪切其邻近位点。结合在左端的单体剪切右端的位点,反之亦然。(我们并不了解这一阶段的单体被激活的反应)。单链的转移反应也是以反式的形式,L1 上的单体转移到 R1 链上;R1 上的单体也转移到 L1 链上。可能不同的单体催化末端的剪切和单链转移反应。

这些反应的产物是一种链转移复合体,在复合体中转座子通过链的每一端和靶位点连接。反应的下一步是不同的,且决定转座的类型。链转移复合体能作为复制的靶子(导致复制性转座)或修复的靶子(导致非复合制性转座)。

五、转座的机制

复制性转座要通过一种**共合体**(图 23－41)。

链转移复合体的 3′端作为复制的引物,这样产生的结构称为共合体(cointegrate),它是由两个复制子融合而成,其中一个含有转座子。共合体含有 2 个拷贝的转座子,位于两个复制子的连接处,方向相同,像是正向重复。这种交换是由转座酶催化的。共合体的形成需宿主的复制功能。

在转座子两个拷贝之间的同源重组释放两个单独的复制子,每个复制子都有一个拷贝的转座子。其中一个复制子是原来的供体复制子,另一个是靶复制子,它通过一个短的 DR 宿主靶序列而获得一个转座子。此重组反应称为**解离**,负责解离的酶叫解离酶。

产生共合体的有关反应的细节在噬菌体 Mu 中得到了解释。1979 年 Shapiro 提出了**转座模型**(图 23－42)这个过程开始是形成**单链转移复合体**(strand transfer complex)(有时也称为**交换复合体** crossover complex),供体和靶的单链连接,这样转座子的两个末端分别与靶及供体单链相连接。连接转移复合体产生了一个交换形成的结构在双链转座子处结合在一起,这个交换结构决定了转座的方式。

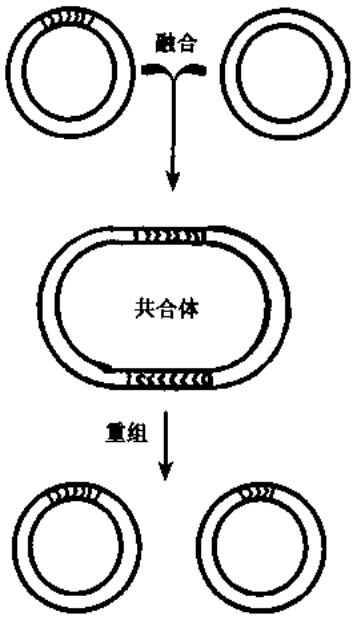


图 23－41　转座子可以使供体和受体复制子融合,形成共合体,解离后释放两个复制子,每个都有一个拷贝的转座子。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18－12)

复制型转座的原理是游离的 3′端以已连结的转座子为模板进行复制,连接,使转座子产生重复,形成共合体。这个模型解释了:① 复制性转座在转座后原来的位置上保留原有的 Tn;② 在新位置上转座子的两端出现正向重复靶序列;③转座过程中出现共合体。

交换结构的每个交错末端含有一个单链区。这个区域是个假的复制叉。它提供了 DNA 合成的模板(这个末端是作为复制的引物,这意味着单链的断裂必需有极性,在此点上产生 3′－OH)。若复制继续形成两个假的复制叉,那么它

的延伸将经过转座子，将它的链分开，在其末端终止复制。复制可能由宿主编码的酶来完成。这个结构变成了共合体。在转座子与复制子之间的连接处有 DR。

六、非复制型转座

交换复合体也可用于非复制型转座，非复制型转座的原理是断裂和重接反应使靶重构。供体保持裂缺。不形成共合体。

图 23－42 表示 MuA 噬菌体的非复制型转座。末断裂的供体单链一旦被切，在转座子另一侧的靶链就能与之连接。通过交错剪切产物产生的单链区必须经修复合成来填补。这个反应的结果是在靶复制子中转座子插在 DR 之间，供体复制子在原转座子位置形成了一个双链缺口。

非复制转座也能经另一途径而产生，这一途径是在靶 DNA 上产生剪切。转座子两侧产生双链断裂，完整的转座子从供体中释放出来（图 23－43）。图 23－43 表示 Tn*10* 的切割和连接的途径。将含有一个错配碱基的 Tn*10* 人工异源双链来进行 Tn*10* 非复

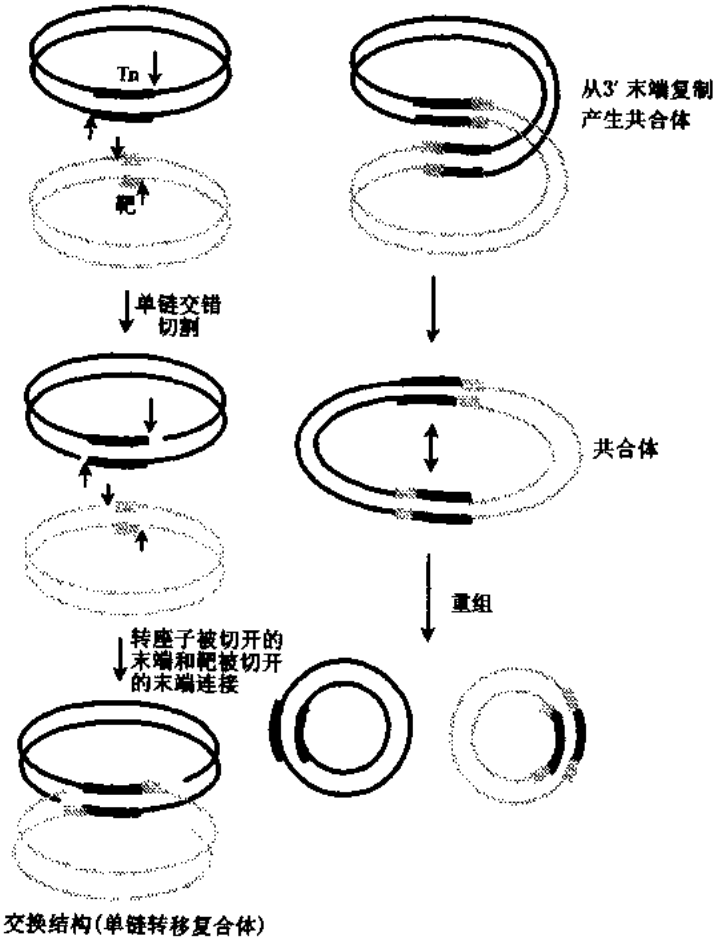


图 23－42　转座模型。Mu 转座产生交换结构，此结构通过复制产生共合体，共合体中转座子产组产生 2 个重组子。（参考 B. Lewin: Genes Ⅵ, 1997）

制性转座试验。若转座涉及复制的话，那么在新位点的转座子应仅含有来自 Tn*10* 亲本的一条链上的信息，若转座是一种被切下的转座子的物理转移，那么在新位点上将会产生基因的转变。错配碱基的转变表明转座子保留了原来两条链的信息。图 23－43 中的模型与图 23－42 中模型的差别在于 Tn*10* 的两条链的剪切发生在与靶序列连接之前。在反应的第一步是转座酶对转座子末端的识别，在转座子的两端的单链按一定顺序被剪切，首先是**转移链**（与靶位点连接的链）被切，然后另一条链被切（与图 23－41 和 42，Mu 噬菌体转座的序列相同）。在此位点释放被切的供体 DNA，转座子与靶位点的切割端相连，转座子和靶位点保持在由转座酶（和其他蛋白）产生的复合体中。在转座子两端的双链剪切排除任何复制型转座，结果产生非复制型转座（图 23－41），但第一次切割和连接按不同序列进行的。Tn*10* 转座酶是一个二聚体，每一个单体都有活性位点，能成功地催化转座子两条链的双链断裂，并交错切割靶位点。

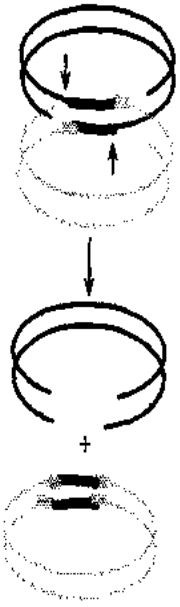


图 23－43 当交换结构剪切释放后导致非复制型转座子插入到靶 DNA 中，DR 包在两侧，供体留下了一个双链缺口。（参考 B. Lewin: Genes Ⅵ, 1997, Fig. 18－14）

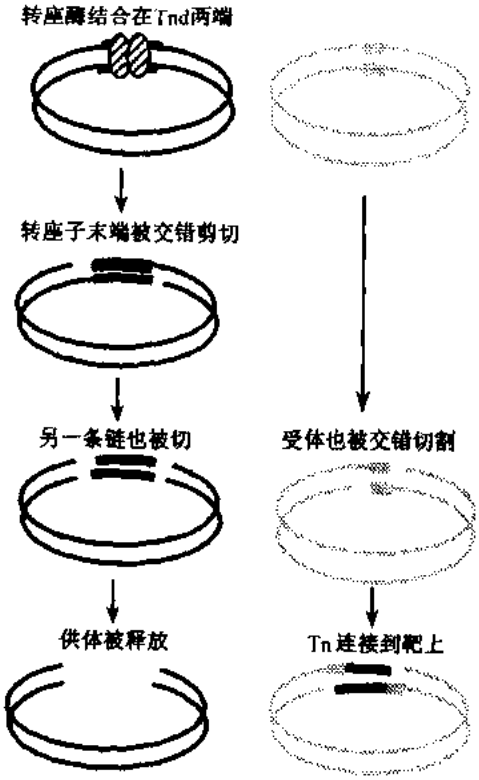


图 23－44 Tn 的两条链先后被切割，然后转座子与切开的靶位点连接。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18－15）

七、Tn*10* 的转座具有多项控制

转座频率的调节对于细胞来说是重要的，一个转座子必须能维持一个最低频率才能存活。若频率太高会损伤宿主细胞。转座子具控制转座频率的机制。在 Tn*10* 中转座频率的调控研究得较为清楚。

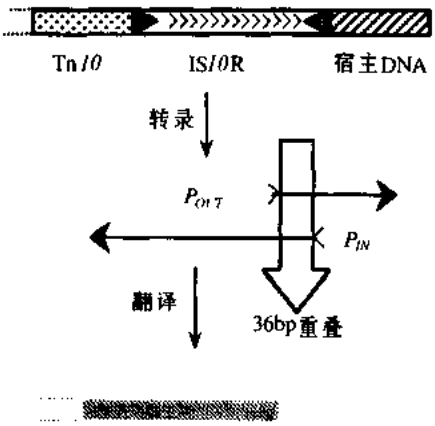


图 23－45　IS10R 中两个方向向反而又部分重叠的启动子。（参考 B.Lewin: Genes Ⅵ, 1997, Fig. 18－17）

Tn*10* 是一种复合转座子，它的右侧 IS*10*R 提供了其转座的活性。IS*10*R 的组成如图 23－44 所示。紧靠分界线处有 2 个启动子。启动子 P_{IN} 是负责 IS*10*R 的转录；启动子 P_{OUT} 负责侧翼相连的宿主 DNA 的转录，转录常在转座子中终止，但有时可持续转录到宿主的 DNA，活化了邻接的细菌的基因。

"多拷贝抑制"（multicopy inhibition）揭示了 IS*10*R 转座的酶基因的表达受到的调节。当将 IS*10*R 的组件装载到多拷贝质粒上，然后将重组质粒导入细菌中，结果在细菌染色体上的 Tn*10* 因子转座能力会减少。这种抑制需要 P_{OUT} 启动子的存在，并以一定的水平进行转录。基本的影响就在于 P_{IN} 转录本的 5′端和 P_{OUT} 的 69 个碱基的转录本有重叠。P_{OUT} 转录的 RNA 比 *IN* RNA 的量要大 100 倍。这有 2 个原因：① P_{OUT} 比 P_{IN} 强得多；② *OUT* RNA 比 *IN* RNA 稳定。

OUT RNA 的功能是作为**反义 RNA**，在其以单拷贝存在时 *OUT* RNA 的水平并不受影响，但当＞ 5 拷贝时就有明显影响。通常是每个拷贝 IS*10* 就有 5 个拷贝的 *OUT* RNA（在典型的状态下相当于约 150 的 *OUT* RNA）。*OUT* RNA 碱基与 *IN* RNA 配对，过量的 *OUT* RNA 保证 *IN* RNA 在和核糖体附着前可迅速地被互补结合，这样一来与 OUT RNA 结合的 *IN* RNA 就不能被翻译。

转座酶的数量是一种关键性的特性。在 Tn*10* 中每个细胞每代只有 0.15 分子的低水平合成，显示出各种有趣的调节机制，图 23－46 表示影响转座频率的几种因素。

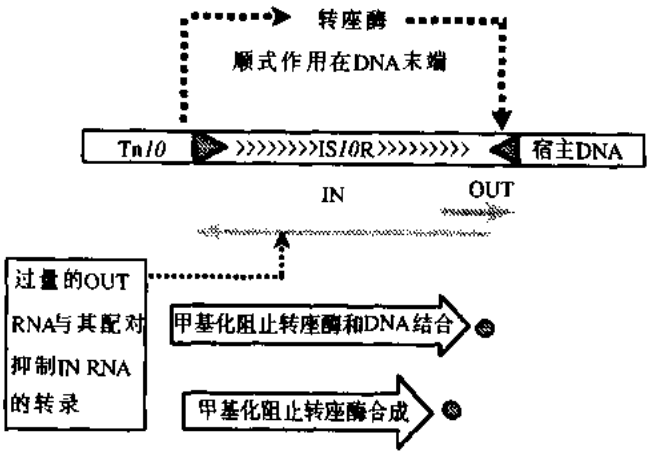


图 23－46　几种抑制 Tn*10* 转座的机制，主要是通过控制转座酶的合成来实现调节。（参考 B.Lewin: Genes Ⅵ, 1997, Fig. 18－18）

在 IS*10*R 的一条链上存在着一个连续的阅读框，它编码转座酶。转座酶的水平限制了转座率。这个基因的突变体能被另一个野生型 IS*10* 因子反式互补，但却不能恢复原有的水平。这反映了转座酶对顺式作用具有很强的倾向。此酶的功能只对带有其转录和翻译的模板链才有效。**顺式优先**（*cis*-prefrence）是 IS 因子编码的转座酶的共同特点。另一些显示顺式优先的蛋白包括和 $\Phi X174$ 复制有关的 A 蛋白。转座酶的顺式优先限制了 Tn*10* 的转座频率，使得在一个细胞中 Tn*10* 多拷贝的存在并不能增加转座酶的浓度，从而保证其转座频率的稳定性。

顺式优先是否反映了转座酶识别 DNA 靶序列的能力比识别酶基因邻近位点更为有效呢？一种可能是转座酶在 DNA 上结合的很紧，以至蛋白合成之后很难扩散到其他地方；另一种可能是这种酶不稳定，它不能固着在 DNA 上，不能结合的蛋白质分子得不到激活的机会。顺式优先和多拷贝抑制的共同作用保证了细菌基因组中 Tn*10* 拷贝数的增加不引起转座频率的增加。

dam 甲基化的作用为单个转座因子提供了重要的调节系统。它们减少转座频率，更重要的是把转座和复制叉的途径联系起来。通过转座子对 2 个位点甲基化的反应体现了 IS*10* 的转座的能力是和复制是相关的。一个位点在 IS*10*R 末端的 IR 中。此是转座酶结合的地方，此位点的甲基化抑制了转录酶和 DNA 的结合。另一个位点是在 P_{IN} 中，这是启动子转录转座酶基因的位点。该位点的甲基化抑制转录合成。

dam 系统可以使以上两个位点甲基化。Dam 甲基化酶可以修饰复制中新合成链上 GATC 序列中的腺嘌呤。在 *dam*$^-$（DNA 腺嘌呤甲基化酶缺失型）系统中，这两个位点未甲基化，而 Tn*10* 转座频率增加了 1000 倍。

复制叉经过这两个位点产生了半甲基化序列；通过转座酶基因从 P_{IN} 更频繁的转录和转座酶对 IS*10* 末端结合能力的增强，转座被激活。在野生型的细菌中复制后的短时间内这两个位点保持半甲基化状态的。

为什么转座在复制后立即发生比较好呢？Tn*10* 转座的非复制机制使供体 DNA 留有空缺（图 23－34），这要冒被"消灭"的危险。若复制正好又产生了另一个供体拷贝，那么就增加了细胞得以幸存的机会。这个机制是有效的，因只有 1 到 2 个新复制的拷贝导致转座（转座子的链在 *dam* 位点末被甲基化）。

由于转座子随机地选择其靶位点，那么就有可能转座子在一个活化的操纵子中着陆。转录从上游开始，延伸时通过转座子，这样激活了转座酶。转座酶生产过剩能否引起高水平的转座呢？Tn*10* 通过两种机制保护自己不受这件事的影响。转录经过 IS*10*R 末端减少它的活性，这大概是通过抑制结合转座酶的能力来实现的。另外从启动子上游延伸的 mRNA 是不太能翻译的，这是因为它有一个二级结构，使起始密码子不易接近。

第七节　真核生物的转座因子

在真核生物中已发现各种不同的转座因子（表 23－6），现分别介绍如下。

一、玉米中的控制因子

早在 1938 年 Marcus Rhoades 分析了一种墨西哥玉米的穗，此穗来自于一种籽粒有颜色的纯种自花受粉的玉米，但它在后代中表现出一种意外的双因子杂种修饰性孟德

表 23 - 6　一些有代表性的真核转座子的属性

物种	类转座子家族	ITR(bp)	TSD(bp)
果蝇	P	31	8
	Hobo	12	8
	Mariner28	2	
	FB	大	9
线虫	Tc1	54	2
玉米	AC/DS	10/11	8
	Spm/dsom	13	3
	Mu/mn	≥200	9
金鱼草	Tam1	13/14	3
	Tam3	12	8/5
一些真核生物	TU/puppy	大	8

ITR:反向重复序列;TSD:靶的返向重复序列

尔分离比(有色：斑点：白色 = 12：3：1),结果表明此现象发生在两个不连锁的座位上,一个座位点控制色素的基因 A_1,它突变为等位基因 a_1,表型为无色素;另一座位的等位基因是 *Dt*(斑点),其表型是在白色的背景上产生有色的斑点。这样原来的品系可能为 A_1 A_1 *dtdt*,突变后产生了 A_1 a_1 *Dtdt* 的植物,自交产生了上述比例(图 23 - 47)。

但是什么因素导致产生花斑呢? 一种可能是在体细胞中产生了回复突变 $a_1 \rightarrow A_1$,但大量的斑点需要很高频率的回复突变。Rhoades 能在 a_1 a_1 *Dt*_(花斑)特殊无性种植物的花中找到相应的花药(图 23 - 48),其花粉应携带回复突变产生色素的基因型,而他用这些花粉与 a_1 a_1 的植株测交,结果有的后代完全是有颜色的。表明在亲本中每个斑点实际上是回复突变的表型效应。这样 a_1 成为首次发现的**不稳定突变等位基因**(unstable mutant allele)的例子。即一种回复突变率很高的等位基因。然而这种等位基因的不稳定取决于不连锁的 *Dt* 基因的存在。一旦回复突变发生,它们就变得稳定了;即 *Dt* 基因能离开 A_1 基因,这时 A_1 表型不再改变。a_1 表型是由一个缺陷型的转座因子

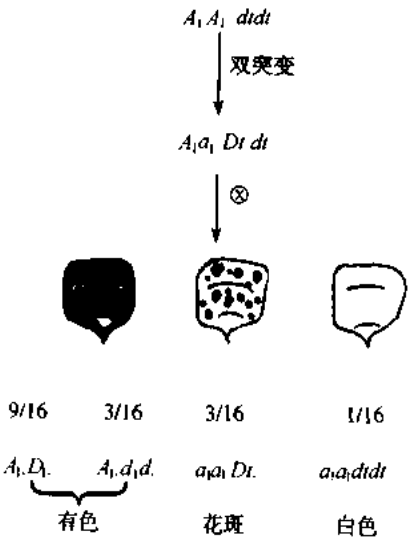


图 23 - 47　玉米花斑表型的遗传学解释。

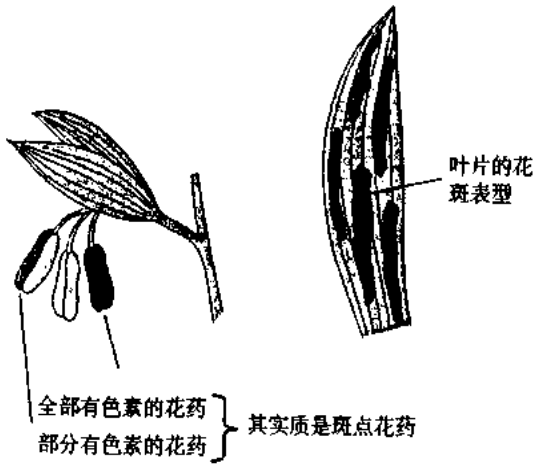


图 23 - 48　Rhoades 将基因型 a_1a_1Dt 的玉米发芽,检测它们是否带有在各组织中产生扇形色素斑的基因。(引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 20 - 33)

的插入而产生的(缺陷的转座因子自己并不能移动)。*Dt* 的缺乏使得表型保持稳定。

1. 麦克林托克的伟大发现

虽然 Rhoades 发现了不稳定基因,而且这种不稳定性是由控制因子所引起,但真正发现控制因子的是 McClintock。她在 1940 和 1950 年间第一个观察了玉米中转座因子。根据她的大量遗传学实验和观察,她描述了大量的"控制因子"(controlling elements),这些因子在玉米中具有修饰和抑制的活性。她研究的性状之一是玉米胚乳上的色素。当时已知很多基因共同控制红色花青素的合成,使玉米胚乳呈紫色。这些基因中任何一对基因发生突变都会影响色素的合成,使胚乳呈白色。McClintock 研究了玉米胚乳的紫色,白色以及白色背景上带有紫色斑点(图 23-49)这些表型的相互关系。她知道花斑表型这种突变是不稳定的。她根据自己的遗传学和细胞学研究推断"花斑"这种表型并不是由常规的突变产生的,而是由于一种控制因子的存在所导致的。我们现在知道这就是转座因子。

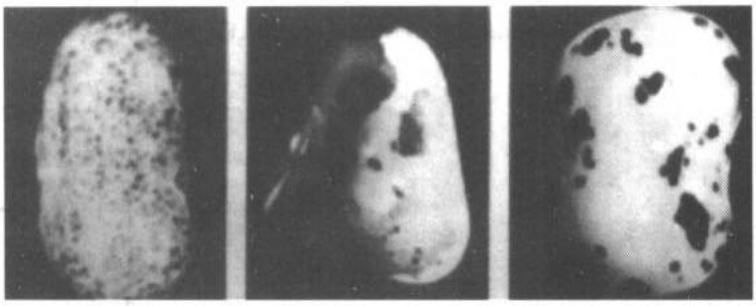


图 23-49 在玉米中由于转座因子转座到控制色素的基因上,使其失活,结果在胚乳上形成了色素斑。(仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 20-12)

她对胚乳的花斑是这样解释的:若玉米带有野生型 *C* 基因,则胚乳呈紫色,*C* 基因的突变阻断了紫色素的合成,那么胚乳呈白色。在胚乳发育的过程中,突变发生回复导致斑点的产生。而回复突变的遗传学性质受到这样一种事实的支持,即细胞的后代经过回复突变也能产生色素。回复突变发生在早期发育阶段,紫斑就比较大。McClintock 认为原来的 *C* 突变(无色素)是由一个"可移动的控制因子"(现在称转座子)引起的,称为 Ds,即解离因子(dissociator)。它可以插入到 *C* 基因中(图 23-50)。我们现在知道这是发生了转座。另一个可移动的控制因子是 *Ac*,称激活因子(activator),它的存在激活 *Ds* 转座进入 *C* 基因或其他基因,也能使 *Ds* 从基因中转出,使突变基因回复,这就是 *Ac-Ds* 系统(图 23-51)。

McClinock 还发现 *Ds* 可导致所在位置的染色体断裂,这种断裂可以通过细胞学的方法和遗传学的方法加以检测。所谓细胞学方法是可以用显微镜进行观察,*Ds* 存在的玉米的 9 号染色体的一条臂上(带有结节,易断裂);遗传学的方法是以一个杂合的品系来观察,当 Ds 存在时,有时其隐性基因可得到表达,表明其显性等位基因发生了缺失所致(图 23-51)。

染色体的断裂结果还会形成**断裂融合桥**(fusion-breakage-bridge)(图 23-52)。*Ds* 的存在可导致其所在位点的染色体断裂,丢失了端粒。经复制后缺少端粒的染色体彼此连结,形成双着丝粒染色体,在有丝分裂中两个着丝点彼此分离向两极移动,形成染色

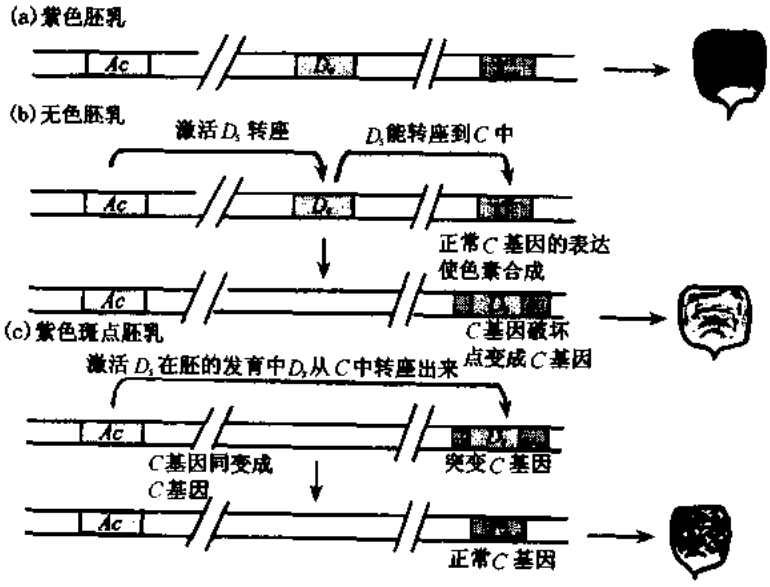


图 23－50　*Ds* 的转座和 *C* 基因的突变和同复突变，引起的花斑表型。
（仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 20－13）

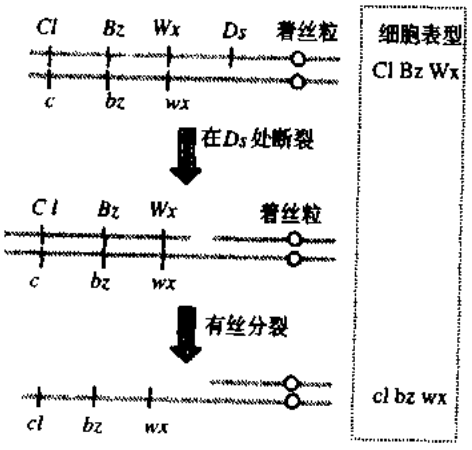


图 23－51　*Ds* 因子导致染色体的断裂，使隐性基因得以表达。（参考 B. Lewin: Genes Ⅵ, 1997, Fig. 18－20）

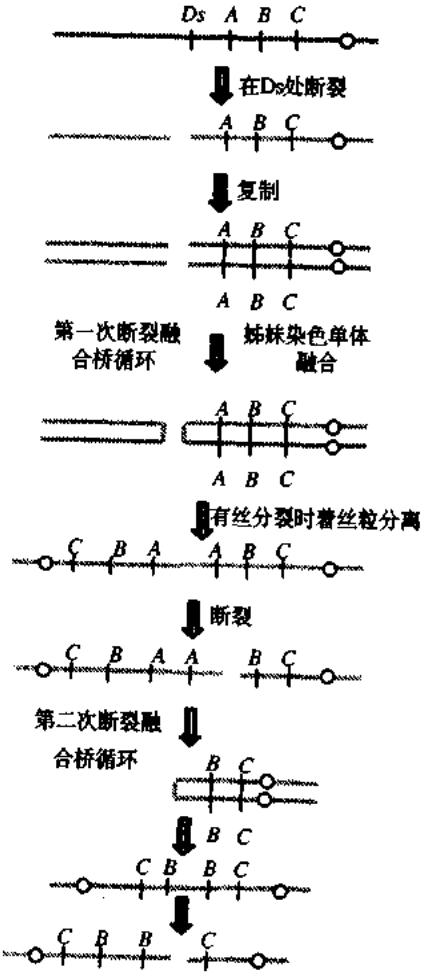


图 23－52　*Ds* 使染色体断裂，形成断裂融合桥。
（参考 B. Lewin: Genes Ⅵ, 1997, Fig. 18－21）

体桥。桥的断裂使一端产生重复，一端产生缺失。这样周而复始形成**融合桥断裂循环**(fusion-bridge-breakage cycle)。

2. 玉米中控制因子家族

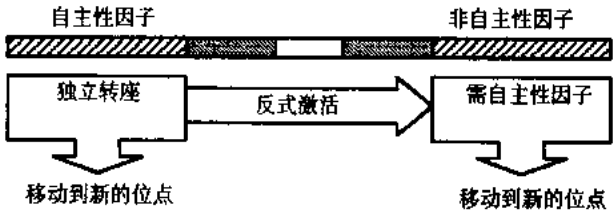
McClintock 不仅发现了 *Ds* - *Ac* 系统，还发现了其他系统(图 23 - 53)，这些因子的成员，类型和位置在各玉米品系中都是特异的。每个家族的成员，可分为两类：

(1)自主性因子(autonomous elements)有切离和转座的能力。由于自主因子的持续活性，它们可以插入到任何位点产生一个不稳定的或可"突变"(mutable)等位基因。自主因子本身的丢失就使可变的等位基因变成稳定的等位基因。

(2)非自主性因子(nonautonomous elements)是稳定的。它们自己并不能转座或自发地改变条件，当同一家族中的不稳定成员存在于基因组中时，不论这自主因子位于何处都能使非自主因子变得不稳定。当通过一个自主因子反式互补时，非自主因子所显示出的一般活性程度和自主因子有关，包括转座到新位点的能力。非自主因子来自于失去了反式作用功能的自主因子，这种功能是转座所必须的。

控制因子家族是通过自主因子与非自主因子的相互作用来划定的。一个家族由单个类型的自主因子组成，这些因子由很多种非自主因子相伴随。一个非自主性因子根据它被什么自主因子反式激活就将它划归这种自主因子的相关家族。在玉米中主要的控制因子家族如图 23 - 53 所示。

玉米转座因子分子水平的特点和原核的转座因子相似。在其两端有反向重复序列



Ac (激活因子) *Mp* (调节因子)	*Ds*(解离因子)
Spm(抑制-突变因子) *En* (加强子)	*DSpm*(缺陷型 *Spm*)
斑点基因(Dotted)	未命名
Mu(增变基因)	不详

图 23 - 53 McClintock 发现在玉米中的转座系统。
(参考 B. Lewin: Genes Ⅵ, 1997, Fig. 18 - 22)

(IR)，在与靶 DNA 连接处有短的正向重复(DR)。但玉米转座因子的大小和编码相差很大。*Ac* 因子和 *Spm* 因子是最有特色的家族。在植物基因组中每个家族有不同的成员(~10)，根据 *Ac*/*Ds* 家族的自主和非自主成员的分析，已掌握了很多成员的结构信息(图 23 - 54)。

在 *Ds* - *Ac* 系统中，大部分自主因子由含 5 个外显子的单个基因组成，其产物是转座酶，它的末端有 11bp 的 IR 和 8bp 的 DR，DR 是由靶位点重复而成。

各种 *Ds* 因子的长度和序列都不相同 (表 23 - 7)，但和 *Ac* 相关。其末端同样有 11bp 的 IR。*Ds* 比 *Ac* 短，其缺失的长度不同，一个极端的例子是 *Ds*9 因子仅缺失 194bp，

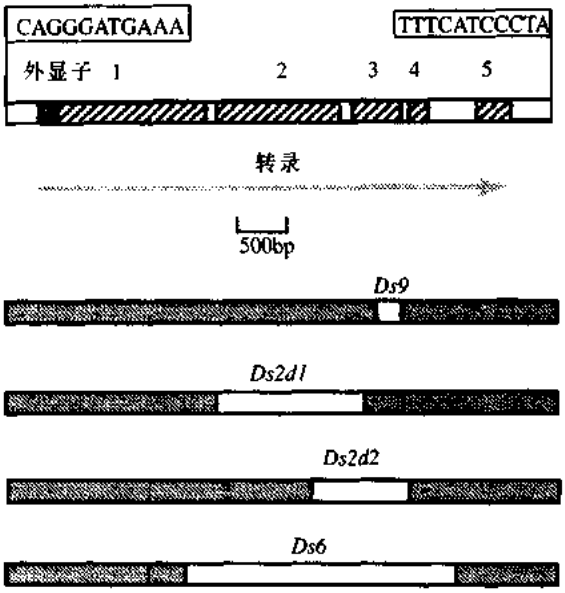


图 23－54　*Ac* 因子有两个开放读框，*Ds* 因子与 *Ac* 的结构相同，但内部有缺失。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18－32）

表 23－7　玉米转座因子的结构

因子	拷贝/基因组	长度(bp)	mRNA(b)	蛋白产物	IR(bp)	靶(DR)
AC	~9	4563	3500	1	11	8
Mul	> 10	1367	?	4 ?	213	9
Spm(*En*)	?	8287	2400	1(+2?)	13	3

B. Lewin: Genes Ⅳ, 1990, Table 33.4

另一个例子是 *Ds*6 因子长仅有 2kb，相当于 *Ac* 两端各 1kb。一个复合重复 *Ds* 因子它有一个 *Ds*6 序列以各种方向插在其他的 *Ds* 中。

非自主因子丢失了内部序列，它具有末端 IR，或其他序列特点。非自主因子来自自主因子，自主因子通过缺失或其他的变化，失去了反式作用的转座酶活性，但留下了完整的转座酶作用位点，包括边界。它们的结构范围从 *Ac* 最小的突变（但失去活性）到有大部分缺失或重排。

在另一个极端 *Ds1* 家族成员含有很短与 *Ac* 末端 IR 相关的序列。这一类因子不需要直接来源于 *Ac*，可以由任何产生 IR 的事件而形成。它们的存在表明转座酶只识别末端的 IR，或者末端重复与某些内部序列相连接。

Ac/*Ds* 发生的转座是通过非复制机制，并且伴随着它们从供体位置的消失。克隆分析表明 *Ac*/*Ds* 总是发生在供体复制之后。这和细菌的转座因子 Tn*10* 的转座位点相似。受体位点常在供体所在的同一条染色体上，而且靠的很近。

复制产生两个拷贝的潜在 *Ac*/*Ds* 的供体，但通常实际上只有一个拷贝的转座酶。转座后供体位点发生了什么情况呢？在供体控制因子丢失位点上出现的重排现象可以用染色体断裂的结果来加以解释（图 23－50）。

自主和非自主因子经历了各种改变，这些变化中有的是遗传的，另一些是外遗传的（epigenetic）。

主要的变化当然是自主因子变成非自主因子，但进一步改变可能发生在非自主因

子中。顺式作用的缺陷可能使非自主因子不受自主因子的影响。这样一个非自主因子可能成为永远稳定的，它变得不再为转座子所激活。

自主因子经受了"**相的改变**"(changes of phase)，其特点是产生可遗传的改变，但这些改变是不稳定的。即在植物的发育中，这些因子以可逆失活的形式在有活性和失活之间产生一种循环。此"相"可影响转座的时间和频率。

在 *AC* 和 *Mu* 型自主因子中的相改变是由于 DNA 甲基化产生的。具有活性和失活的控制因子对限制性酶的易感性表明这些因子的失活是由于在其靶序列 $\begin{matrix}\text{CAG}\\ \text{GTC}^{*}\end{matrix}$ 位点被甲基化了。每个因子具有各种靶位点，我们还不知道这些靶位点的控制作用。人们希望搞清楚是什么在控制此元件的甲基化和去甲基化。

可能存在着转座的自我调节控制，和细菌转座子所显示的免疫作用相类似。在基因组中 *Ac* 因子数量的增加会减少转座频率。*Ac* 因子可能编码转座的抑制物，具有转座功能的某些蛋白可能带有这种活性。

Spm 和 *En* 自主因子实际上是相同的。它们仅在不到 10 个位置上有差异(图 23-55)。就像其他的转座子一样，末端含有 13bp 的 IR，此重复序列对于转座是必须的，末端缺失就会形成转座的缺陷型。与 *Spm* 相关的转座子在其他的植物中也有发现，它们的结构相似，属于同一个家族。它们末端 IR 都邻接着靶 DNA 重复产生 3bp 的 DR。末端的 IR 称为转座的 CACTA 群。

一个 8.3kb 的序列是由这个基因左端的启动子转录而来。转录本中含 11 个外显子接成 2.5kb mRNA。它编码 621 个氨基酸的蛋白，这个基因称为 *tnpA*，这个蛋白结合在此因子末端区域中的多拷贝存在的 12bp 的保守序列上。*TnpA* 的功能是剪切，但可能还需其他因子的参与。

这个家族所有的非自主因子(*dSpm* 是缺陷的 *Spm*)都和 *Spm* 因子本身的结构密切相关。它们是缺失了外显子的 *tnpA*。

两个附加开放阅读框 (ORF1 和 ORF2) 位于 *tnpA* 的第一个长的内含子中，它们含

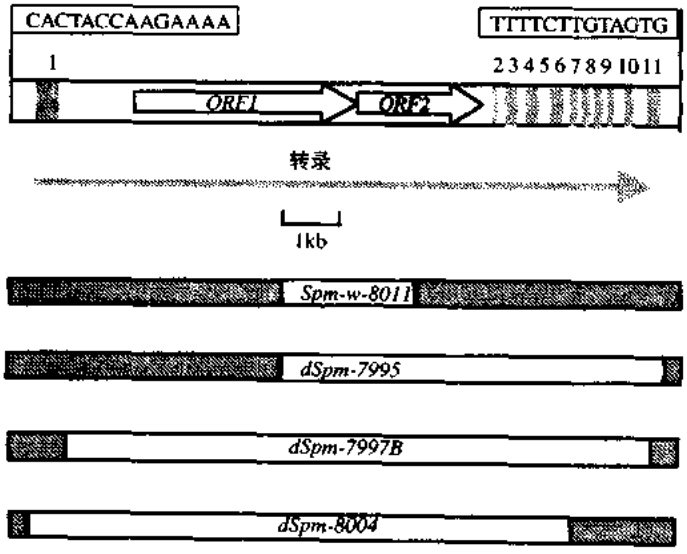


图 23-55 *Spm/En* 有两个基因，*tnpA* 有 11 个外显子转录成 2500b 拼接 mRNA，*tnpB* 含有 6kb mRNA，含 2 个读框。(仿 B. Lewin: Genes VI, 1997, Fig. 18-24)

有选择性连接的 6000b 的 RNA。在 *tnpA* 的 *mRNA* 中以 1%的水平存在这种 RNA。含有 ORF1 和 ORF2 的因子称为 *tnpB*。它产生的蛋白可能结合在 13bp 的末端 IR 上，剪切末端进行转座。

除了完整的激活的 *Spm* 因子外，还有 *Spm*-w 可诱导较弱的转座活性。它是缺失了 ORF1 和 ORF2 产生的（图 23－55），这表明在转座中，*TnpB* 要通过其他途径的作用或被取代才行。

Spm 的插入能控制靶位点基因的表达，受体位点可能受到正的或负的调控。*Spm*-可抑制基因座（*Spm*-suppressible locus）受到抑制而不能表达。*spm*-依赖性座位（*spm*-dependent locus）只有在 *spm* 的帮助下才表达。当被插入的因子是一个 *dSpm* 时，对转移功能产生抑制或依赖自主性 *Spm* 来提供帮助。这两个相反作用的基础是什么呢？

一个 *dSpm*- 可抑制等位基因中其外显子内插入了一个 *dSpm*，人们对这个结构会立即产生疑问，一个基因其外显子中插入了一个 *dSpm* 它怎么能表达！其实这个 *dSpm* 序列能在转录本中利用这个序列的末端被剪切掉。此剪切事件可以使 mRNA 序列中留下一个变化。这样，就解释了它所编码的蛋白性质发生改变的原因。同样某些插入的 *Ds* 也可从转录本中被剪切掉。

TnpA 原称为 *Spm* 因子，它提供了抑制功能。缺陷型因子的存在可能使它所插入的基因表达减少，但并不消失。然而诱导有功能的 t*npA* 基因的自发因子可能会抑制靶基因的表达。抑制作用是 *TnpA* 结合于缺陷因子中的靶位点的能力产生的，从而阻断了正在进行的转录。

一个 *dSpm*⁻依赖性等位基因在其附近（而不在基因中）含有一个插入序列。这个插入序列提供了一个增强子，它可以激活位于受体座位的基因的启动子。

在 *dSpm* 因子上的抑制和依赖的存在取决于一个自主因子 *Spm* 因子 *tnpA* 基因的反式作用产物与这个因子末端的顺式作用位点间的相互作用。因此在蛋白质和此因子末端之间的单个相互作用不是抑制就是激活受体基因上游或者受体基因中的靶座位。无论靶位点否依赖这种因子。

Spm 因子从完全活化到隐蔽以各种状态存在。隐蔽因子是沉默的，既不转座也不激活 *dspm* 因子。一个潜伏的因子可以通过和完全活化的 *Spm* 因子的相互作用而转变或恢复成活性状态。失活是由于在转录起始附近的序列被甲基化而导致的。一个因子被重新甲基化而失活或经过去甲基化而恢复活性这一现象的本质尚不清楚。

二、果蝇中的 P 因子

在某些果蝇品系中杂种繁殖遇到困难，当将这种品系的果蝇进行杂交产生的二倍体是“杂种不育”的品系。它们含有一系列缺陷，包括突变，染色体畸变，减数分裂的异常分离和不育。

在果蝇中已鉴别出两个系统与杂种不育有关。在第一个系统中，**I**(inducer)型雄果蝇和 **R**(reactive)**型**雌果蝇杂交会降低生育力，但反交并不如此。第二个系统的果蝇分成为 **P 型**（父本贡献的，paternal contributing）和 **M 型**（母本贡献的，matarnal contributing）。图 23－56 表示 M(♂)×P(♀)后代可育，但反交 P(♂)×M(♀)后代不育。

不育原则上是配子细胞的一种现象。及 P－M 系统的杂交中，F_1 代具有正常的体细胞组织，在生殖细胞迅速开始分裂时，它们的性腺并不能发育。

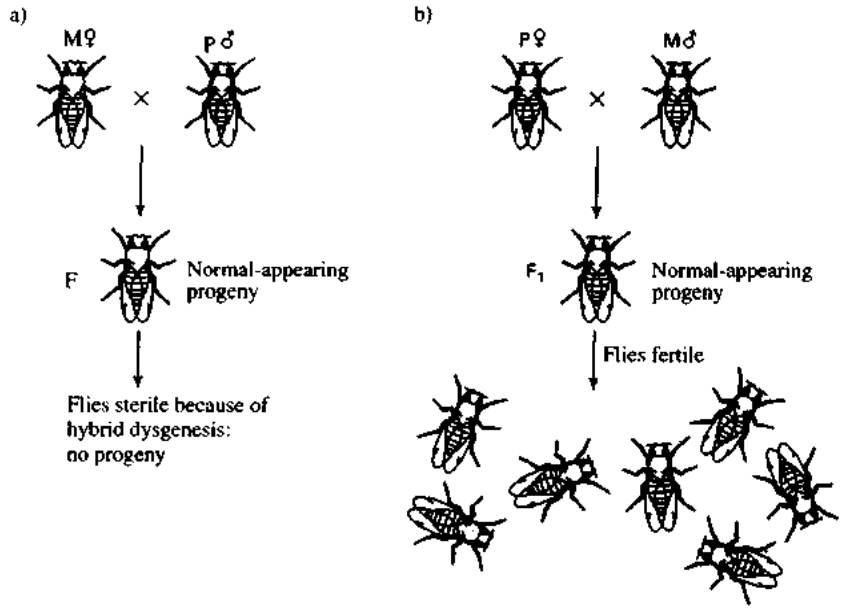


图 23－56 杂种不育。(a)M♀ × P♂ 产生的杂种后代不育;(b)P♀ × M♂ 产生的杂种后代可育。(仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 20－19)

在和 M 雌果蝇杂交中 P 型雄果蝇的任何一条染色体都能导致不育。染色体重组的实验表明,在每条 P 型雄果蝇染色体中的各区域也都导致不育。这表明 P 型雄果蝇具有大量的 **P 因子**(P elements),这个因子存在于很多不同的染色体位置上。这些位置在 P 品系的个体之间也是不同的。而在 M 雌果蝇的染色体上都没有这种 P 因子。

通过对杂种不育果蝇的 W 突变体的 DNA 作图发现，不育是 P 因子的存在导致的。所有的突变都是由于 W 位点插入了 DNA 片段。这个插入序列被称为 **P 因子**（P element)。P 因子的插入形成了一种可转座系统。各个 P 因子长度不同,但序列同源。所有的 P 因子末端都具有 31bp 的反向重复。而且使靶序列产生 8 个 bp 的 DR。最长的 P 因子长 ~ 2.9kb,有 4 个开放阅读框。较短的 P 因子经常出现,此是由于其内部缺失而形成的。在某些较短的 P 因子中至少失去了产生转座酶的能力,但可以被完整 P 因子编码的酶反式激活(图 23－56)。

一个 P 品系果蝇带有 30 ~ 50 拷贝的 P 因子,其中约 1/3 具有完整的结构。M 品系失去了这种因子。在一个 P 品系中,这些因子是作为基因组中没有作用的成分。但当一个 P 雄果蝇和 M 雌性杂交时它们被激活转座。

在 P－M 杂种不育果蝇的染色体中，P 因子插入到很多新的位点。典型的杂种不育果蝇染色体在热点发生断裂,所谓热点是对 P 因子敏感位点。P 因子对 M 染色体的平均转座率约是每代一次。

P 因子的激活是组织特异性的:它仅发生在生殖细胞中。但 P 因子在体细胞和生殖中都转录。组织特异性是表现在不同的细胞中转录的方式不同。

图 23－57 表示 P 因子的结构和转录。初始转录本有 2.5kb 和 3.0kb 两种不同的长度,差别在于一种缺少了终止位点。不同的组织中 P 因子可以产生两种蛋白(图 23－57)。① 在体细胞组织中,仅前 2 个内含子被剪切。产生了 ORF0－ORF1－ORF2 编码区。它翻译成 66kDa 的蛋白,这种蛋白是转座激活的阻遏物。②在生殖系统中,3 个都被剪切掉,4 个阅读框全部拼接在一起形成 mRNA,它翻译产生一个 87kDa 的蛋白。此

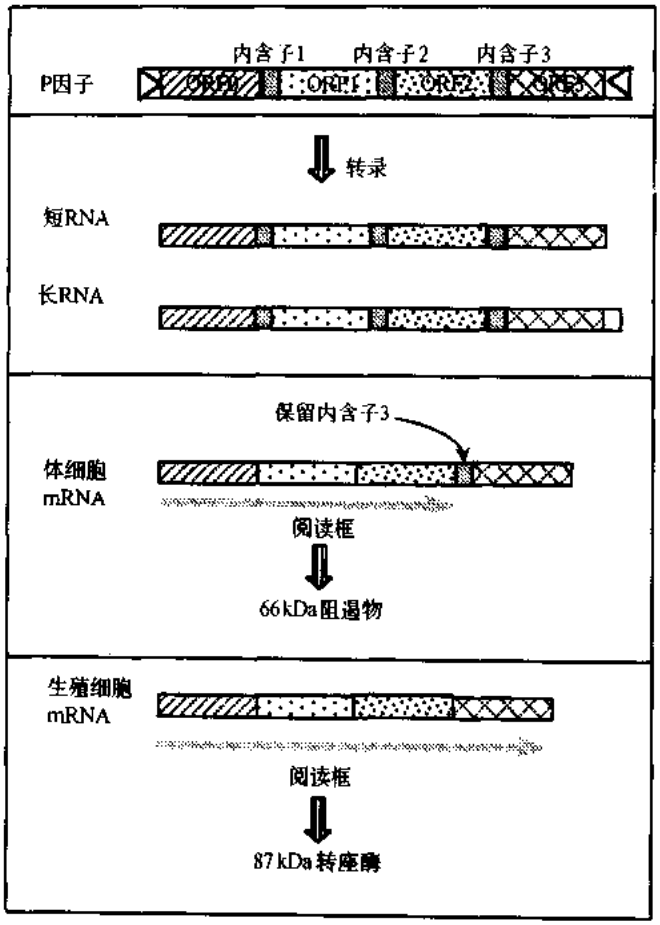


图 23－57　P 因子有 4 个外显子。在体细胞中前 3 个拼接在一起表达；在生殖细胞中 4 个外显子拼接在一起表达。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18－26）

蛋白是转座酶。

两种类型的表达证实了第 3 个内含子的剪接对转座是必要的。首先，若剪切连接在体外发生突变，并将 P 因子导入果蝇则不能完成转座激活；第二若第 3 个内含子缺失。ORF3 在各组织中都会组织到 mRNA 中，那么体细胞中也会和生殖细胞中一样发生转座。

无论何时 ORF3 被拼接成阅读框，P 因子就成为有活性的。这是一个关键的调控作用，通常它仅发生在生殖细胞中。

是什么负责组织的特异拼接呢？体细胞中含有一种蛋白，它结合在第 3 个外显子上，阻止这个内含子的剪接。在生殖细胞中缺乏这个蛋白，因此可以剪接第三个内含子产生编码转座酶的 mRNA。

P 因子的转座需 150bp 左右的末端 DNA。转座酶结合在 10bp 的序列上，这序列和 31bp 的 IR 邻接。通过一种 Tn*10* 式的非复制型“切割和粘贴”机制发生转座。（它以两种方式产生杂种不育。转座因子在新位置的插入可导致突变。在供体位点留下的断裂（图 23－34）具有不利的影响。

有趣的是一定比例的供体 DNA 缺口被同源序列所修复。若同源物有一个 P 因子，那么一个 P 因子存在于供体位点中可能是被修复时产生的。若同源物中无 P 因子，那么修复后的供体位点也没有 P 因子。这样明显产生了一个精确的删除。

杂种不育依赖于杂交的性别的定向，表明细胞质和 P 因子同样重要。细胞质的贡献被描述为**细胞型**(cytotype)：在含有 P 因子的品系中有 **P 细胞型**，而缺少 P 因子果蝇品系具有 **M 细胞型**。杂种不育仅发生含有 P 因子的染色体进入 M 细胞型时，即雄性亲本有 P 因子，而雌性亲本却没有 P 因子的情况。

细胞型表明一个可遗传的细胞质因子。P 细胞型(雌性亲本有 P 因子)M 雄性亲本杂交产生的后代来说杂种不育被抑制了(图 23－58)。

细胞型的作用如图 23－58 所示。取决于 66kDa 蛋白抑制转座的能力。这个蛋白质是卵中母体因子产生的。在 P 品系中必须要有足够的蛋白来抑制转座的发生。含有 P 因子的雌果蝇与无论是否带有 P 因子的雄果蝇杂交，由于 P 细胞型的存在抑制了转座酶的合成或激活。但当雌果蝇为 M 型时，在卵中无阻遏物，这样导致了来自雄果蝇中的 P 因子使生殖细胞中转座酶活化。P 细胞型能对多代后代发挥作用，表明卵中必须要有足够的阻遏蛋白，而且也要相当稳定，通过成体传递到下一代的卵中。

30 年前在野外捕到的黑腹果蝇几乎都是 M 品系，而最近 10 年捕到的几乎都是 P 品系。这是否意味着在最近几年 P 因子家族侵入到果蝇的群体中了呢？P 因子当它导入到一个新的群体之中便已高度侵入了；侵入因子可能起源于另一物种。由于杂种不育减少了品种间的杂交，是新种形成途径的一个步骤。

推测杂种不育系统是在某些地理位置通过转座产生的。另一些因子可能是在其他

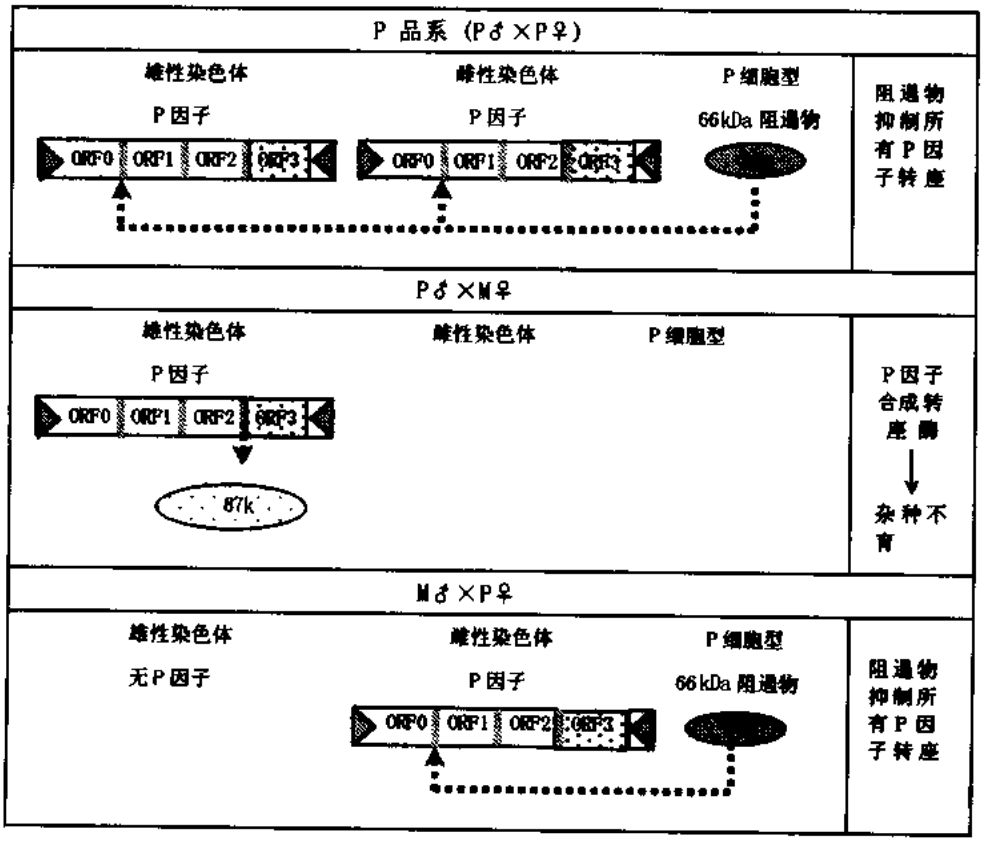


图 23－58　杂种不育取决于基因组中 P 因子和不同类型细胞中 66kDa 阻遏物的相互作用。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18－27)

某些地理位置产生的不同系统。两个不同区域的果蝇因是两个不同的系统而将产生杂种不育。若这一表现使它们之间杂种不育那么群体将出现隔离，进一步的隔离可能还会发生，多个杂种不育系统导致它们之间不能交配而形成新种。

第八节　反转录病毒和反转录子

一、反转录病毒

1. 反转录病毒的生活史

由 RNA 介导的转座仅发生在真核生物中，它是由**反转录病毒**（retroviruses）以其 RNA 基因组的 DNA 拷贝插入到宿主细胞的染色体中而产生的。有些生物的转座子和它们组织中的反转录病毒的前病毒有关。而且它们的转座是通过 RNA 介导的。这一类因子称为**反转录子**（retroposons），有时称为**反转录转座子**（retrotransposons）。它们的范围从可以自由感染宿主细胞的反转录病毒本身到利用 RNA 转录的序列，但这些序列本身并不能转座。它们和其他类型的转座子相同，具有转座子的重要特征：在插入位点的靶 DNA 上产生正向重复。但在基因组中难以检测到活化的转座子，转座事件的踪迹是在其侧翼分布的重复序列中发现了正向的靶重复。这些序列的特点有时表明一个 RNA 序列作为 DNA 基因组序列的祖先。RNA 必须转变成 DNA 拷贝，通过类似转座的行为才能插入到基因组中。

和其他的繁殖循环一样反转录病毒或反转录子的循环也是连续的，反转录具有单链 RNA 基因组，它通过反转录酶以 RNA 为模板合成双链 DNA，然后整合到宿主的基因组中，使靶 DNA 产生了正向重复序列。

反转录病毒整合在细胞的基因组中，作为一**种内生病毒**（endogenous provirus），像是一个溶源性细菌噬菌体，其行为表现为生物遗传物质的一部分。

细胞的 DNA 与反转录病毒序列重组，然后与其一起转座，这些序列可能被整合到基因组中，在新的位点中反转录病毒序列以双链形式存在。

被转座的细胞序列可能改变了细胞的特点，图 23－59 表明反转录病毒的特殊生活周期。关键的步骤是病毒的 RNA 反转录成 DNA 整合到宿主的基因组中，然后 DNA 又转录成 RNA。

负责将病毒 RNA 转录成 DNA 拷贝的是反转录酶（reverse transcriptase）。在病毒感染的细胞质中，它以 RNA 为模板合成双链 DNA，此 DNA 也可能形成环状分子，但环状分子和病毒的增殖无关。它并不能整合到靶 DNA 中。

线性的 DNA 能进入核中，其一个或多个拷贝整合到宿主的基因组中，整合酶（integrase）负责整合。整合进入宿主基因组的前病毒 DNA（proviral DNA）由宿主的转录酶系转录成病毒的 RNAs，此 RNAs 有两个用途：作为翻译模板的 mRNA 和包装成病毒的基因组。整合是其生活周期中正常的步骤也是进一转录所必要的。

两个拷贝的 RNA 基因组被包装进一个病毒颗粒中，使每个病毒都成了特殊的二倍体。当一个细胞同时感染两个不同的病毒时，可能产生一种杂合的病毒颗粒，它带有的基因组是 2 种类型的 RNA。这种二倍性对于获得细胞的序列是非常必要的。在病毒颗粒的基因组中携带着反转录酶和整合酶。

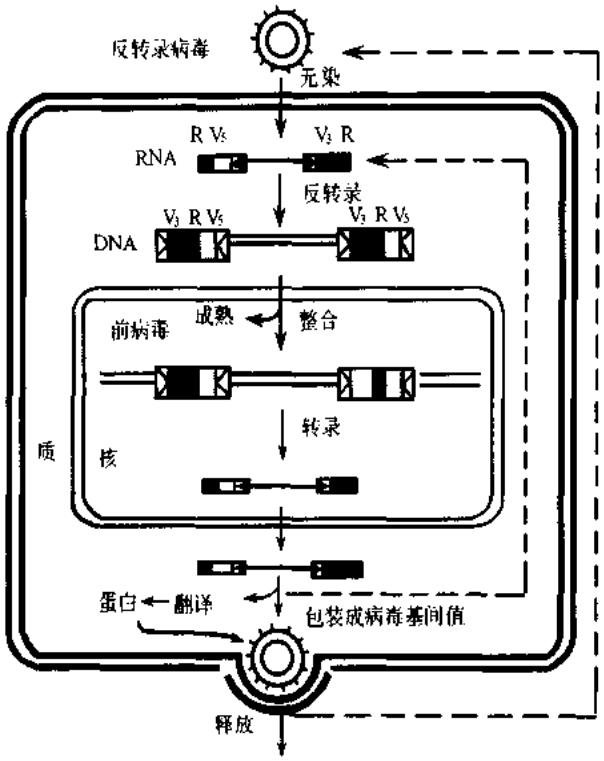


图 23－59　反转录病毒的繁殖周期。

2. 反转录病毒基因组的结构与功能

典型的反转录病毒含有 3～4 个"基因"，是由 *gag* – *pol* – *env* 组成的(图 23－60)。每一个基因所编码的肽链经加工后可产生多种蛋白(图 23－61)。

反转录病毒 mRNA 具有常规的结构：在 5′端有帽结构，在 3′端有多聚腺苷的尾巴。它存在两种 mRNA：长的 mRNA 翻译 Gag 和 pol 多蛋白。从起始密码子到第一个终止密码的阅读框翻译成 Gag 产物。pol 表达一定要越过第一个终止密码子。

不同的病毒采用不同的机制来越过第一个终止密码子，它取决于 *gag* 和 *pol* 两个阅读框之间的关系。当 *gag* 和 *pol* 需要连读的时候，通过可识别此终止密码子的 Glu－tRNA 来抑制终止，产生单个的蛋白。当 *gag* 和 *pol* 在不同读框中，核糖体发生移框来产生单个的蛋白。通常连读率约在 5%，因此 Gag 蛋白在数量上胜过 Gag－pol 蛋白约 20 倍。

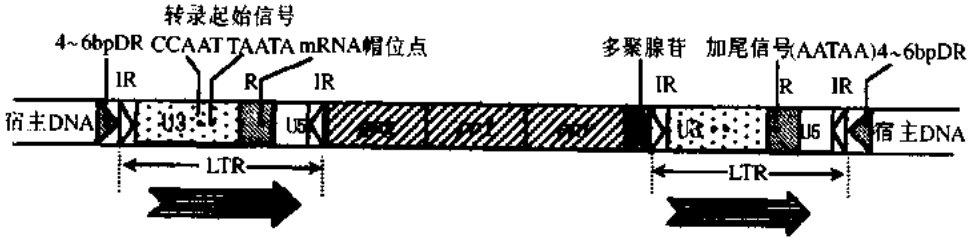


图 23－60　LTRs 的特点。大部分反转录病毒和反转录转座子都有 LTR。
(仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 20－22)

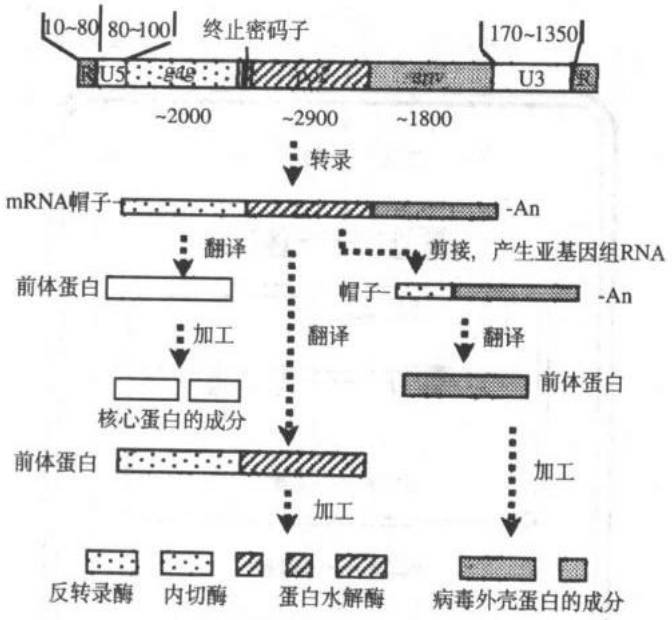


图 23－61　前病毒的转录，翻译和加工，产生多种蛋白。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 19－3)

Env 多蛋白通过另一种机制进行表达，它是将 mRNA 重新拼接成一个短的亚基因组的(subgenomic)信使，翻译成 Env 产物(图 23－61)。

Gag 或 Gag－pol 和 Env 产生的是**多蛋白**(polyprotein)，它能被蛋白水解酶切成多种单个的蛋白，存在于成熟的的病毒中，这种蛋白酶的活性是由病毒编码的。它可能是 gag 或 pol 的一部分，有时又可以一个独立的附加读框存在。

反转录颗粒的产物涉及到将 RNA 包装进核心蛋白中，外周包以衣壳蛋白和从宿主细胞膜上取下的片段。图 23－62 表示病毒颗粒从细胞中释放的过程。感染正是这一过程的逆向反应。病毒是通过和一个新的宿主细胞质膜的融合来感染的，然后将病毒粒子的内容物释放到宿主细胞中。

3. 反转录病毒的整合模型

反转录病毒被称为正链病毒(plus strand viruses)，这是由于病毒的 RNA 本身编码了蛋白产物。反转酶负责将基因变成为互补的 DNA 链，这链称为负链 DNA(minus strand DNA)。反转录酶在产生双链的阶段催化底物，它具有 DNA 聚合酶的活性，它可以从病毒的单链 RNA 转录本来合成双链 DNA。双链中的第二条 DNA 单链称为正链 DNA(plus strand DNA)。反转录酶还具有 RNase H 的活性，可以降解 RNA－DNA 杂种链中的 RNA 部分。各种反转录病毒的反转录酶中相似的氨基酸序列是很重要的，这些

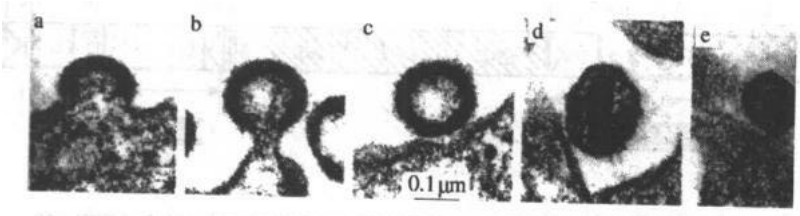


图 23－62　反转录病毒从被感染细胞的细胞膜上突出。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 19－4)

同源序列可能被某些反转录转座子所识别。

来自病毒的DNA链的结构与RNA互补(图23-63),在其RNA的两端具有正向重复序列。在不同的病毒中此DR区-R序列差异很大,长10到80nt不等。在5′端紧跟着R的是U5区,长80~100nt,其命名表示它们于5′端;在3′端R的内侧是U3序列,长

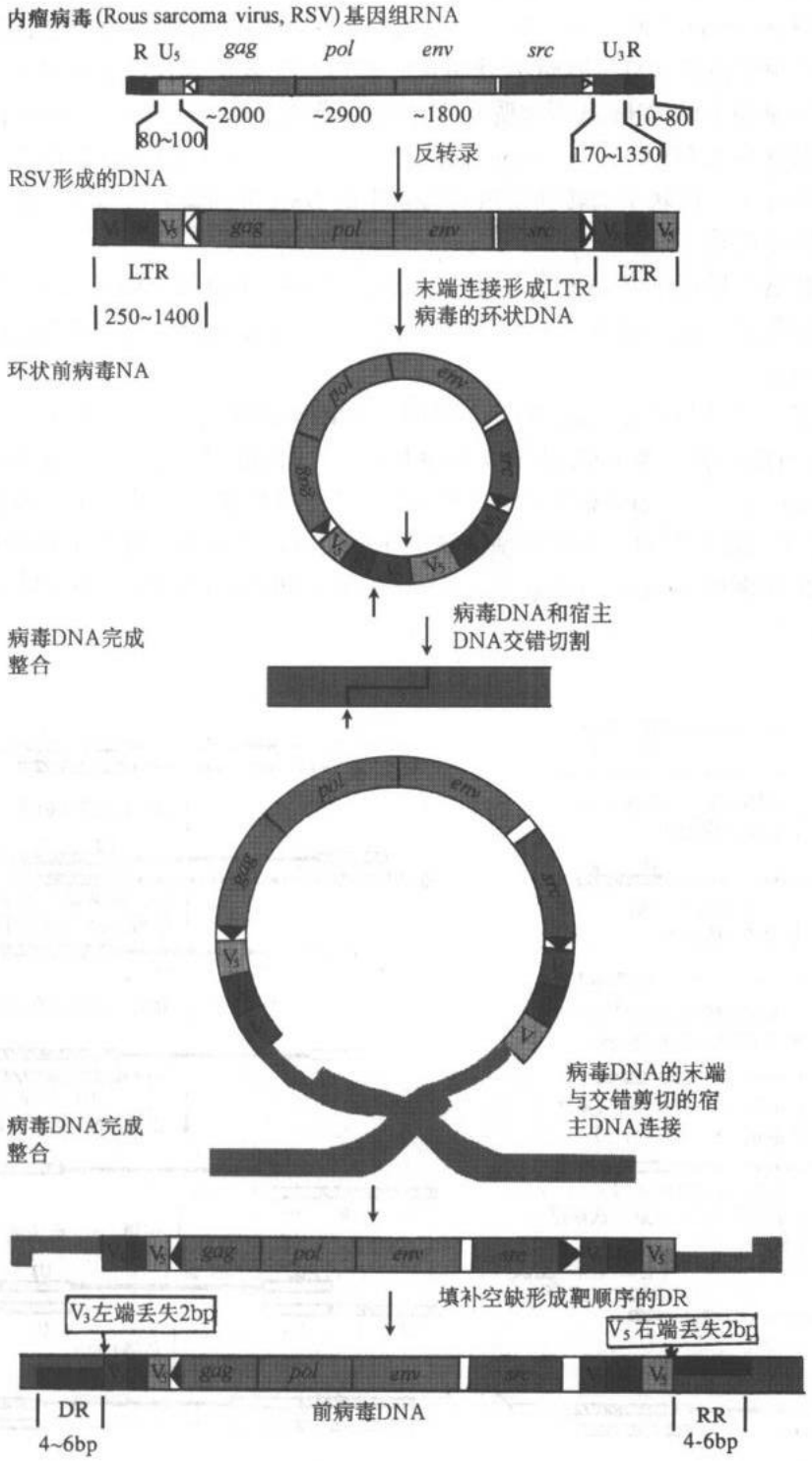


图23-63 RSV RNA基因组以及其整合列宿主的染色体中。

(仿 Peter. J Russell: Genetics, 3rd ed., 1992, Fig. 20-21)

170～135nt，它位于3′端。R片段在病毒RNA变成DNA时用于和另一条基因组RNA配对(图23－64)。

像其他的DNA聚合酶一样，反转录酶也需要一个引物。天然的引物是tRNA，此tRNA原来就存在于病毒中。tRNA的3′端有18nt的序列与病毒RNA分子端100～200碱基处的位点(PB)配对，也可与病毒的另一条RNA 5′端的PB′位点配对这样有助于在病毒的RNAs之间形成二聚体。这儿有一个看似矛盾的地方，就是反转录酶是在RNAs5′端下游仅100～200nt处开始合成DNA，那么按其合成方向是向左的，这样怎能反转录整个的RNA基因组呢？在体外合成到末端会产生一个短的DNA序列，称为**负链强停止DNA**(minus strong-stop DNA)，在体内合成是连续的(图23－63)，故未发现这种分子，反转录酶转换模板时将新生的DNA带到新的模板上，这是在模板间二次跳跃中的第一跳跃。

在这个反应中RNA模板的5′端的R区被反转录酶的RNase H活性所降解，R区的切除随着与新合成的DNA 3′端配对进行的。然后反转录酶持续通过U3进入RNA的编码区。

进一步与强停止负链DNA配对的R区的来源可能是同一条RNA分子3′端的R区，也可能是另一条RNA分子3′端的R区(分子间配对)。这种不同RNA模板的转换如图23－63。这是因为有证据表明tRNA引物序列在反转录子生活周期中并不遗传(若分子间发生配对，我们将预期这个序列是可以遗传的，因在下轮循环中它仅提供**引物结合序列**(primer binding sequence，PB)分子间的配对使另一条反转录RNA也提供PB序列。

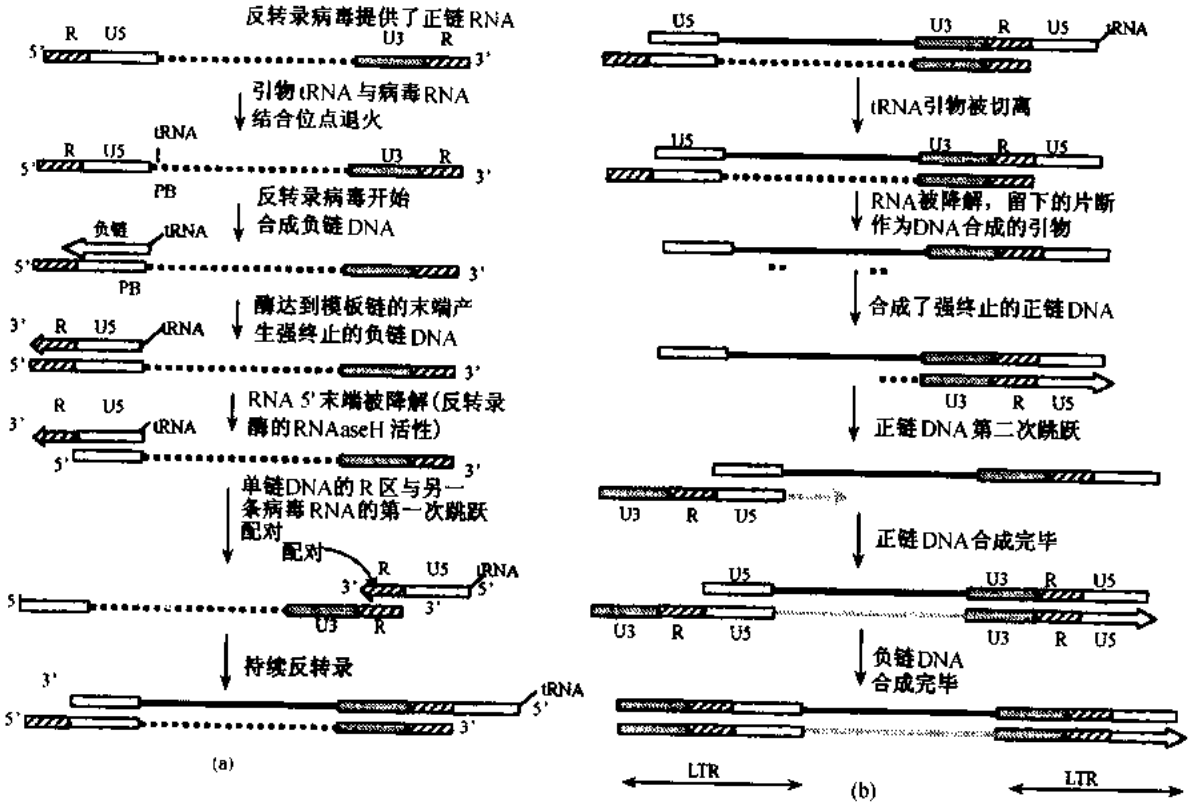


图23－64　反转录病毒的RNA反转录成DNA双链。(a)由模板转换产生负链DNA。(b)通过第二次跳跃合成正链DNA。(仿B. Lewin: Genes Ⅵ, 1997, Fig. 6－7)

模板的转换和延伸的结果是将 U3 序列加到 5′端同时也将 US 加到了 3′端，使两端产生了相同的序列。这种序列结构“U3－R－U5”被称为长的末端重复序列（long terminal repeat ，LTR）。

我们需要产生一条 DNA 正链，每端都形成 LTR1。此反应如图 23－64(b)所示，反转录酶以降解余下的 RNA 为引物合成正链 DNA，当酶到达模板的末端就产生了一个强停止正链 DNA。这个 DNA 再移到负链 DNA 的另一端。有可能当第二轮 DNA 合成从引物片段开始，上游发生取代反应将它释放出来。它利用 R 区域与负链 DNA 的 3′端配对。然后双链 DNA 的形成需要两条链的每一末端产生双链的 LTR 才能完成。

由于在每个反转录病毒颗粒中有两条基因组 mRNA。可能在病毒生活周期中发生重组，此和负链及正链的合成有关。① 在分子间配对时（图 23－64a）在两个连续的 RNA 分子间发生重组。②正链 DNA 的合成是不连续的，在反应中涉及到各种内部的起始。在这个反应中链也可能发生转移。

这两件事的共同特点是在 DNA 合成时重组是由模板的转换而产生的。这种重组的机制称为**拷贝选择**(copy choice)，前面已提到在多年前解释常规重组时，人们曾提出这一模型。但它并不能应用于细胞系统。但在 RNA 病毒感染时它却是重组的普遍基础，包括独特的通过 RNA 形式复制的病毒，例如脊髓灰质炎病毒。

含有前病毒的生物都具有线性 DNA，病毒两端的 LTRs 是相同的。US 的 3′端和 U3 的 5′端由一个短的 IR 组成的（图 23－59），因此 LTR 本身的两个末端都有反向重复序列，而它的侧翼是靶 DNA 形成短的正向重复序列（图 23－59）。整合的前病毒 DNA 结构类似转座子。

前病毒是通过将线性 DNA 正向插入到靶位点中产生的。另外线性 DNA 也形成环型的 DNA 序列。两个 LTR 序列连接起来产生一个线性的末端。另一种情况是只有一个 LTR，据推测可能是通过重组事件，实际上是由环组成的。虽然长时间都认为中间分子是一个环（由此是按照 λ 噬菌体 DNA 的整合的模型推测的），我们现在知道整合时是用线性的 DNA，而不是环。

线性 DNA 的体外整合已被描叙，并由病毒的整合酶所催化。整合酶可以作用于反转录病毒的线性 DNA 和靶序列，此反应如图 23－65 所示。

病毒 DNA 的末端是重要的；就像转座子中的情况相似，若发生突变就会抑制整合。其最保守的特点是在每个反向重复序列的末端有 2 nt 的序列 CA，整合酶就结合在这个线性 DNA 的末端，形成 RNP 的形式，并将两个平端的末端通过切除 CA 2 个碱基变成为有缺口(缺少 2 个碱基的)3′末端。

靶位点的选择是随机的，整合酶在位点上交错切割，如图 23－65 所示，切点相隔 4 个bp，这个长度就是整合后前病毒两侧正向重复的长度，不同的病毒此长度不同。可以是 4、5 或者 6 个 bp，可能这是由整合酶作用的病毒基因组与靶 DNA 之间的反应决定的。

靶 DNA 经剪切后产生的 5′末端与病毒空缺的 3′端共价连接。在这一位点病毒 DNA 的两个末端由一条单链与靶 DNA 连接。单链区域经宿主细胞中的酶来修复。在这个反应过程中病毒 DNA 5′端 2 个碱基的凸出被切除掉，结果使整合的病毒 DNA 两端的 LTR 丢失了 2bp，而在整合的反转录病毒基因组的两端产生了靶 DNA 的短重复序列。

病毒的 DNA 在随机选择位点上整合到宿主的基因组中，一个连续感染的细胞可获得 1～10 个拷贝的前病毒。一个感染的病毒进入细胞质中，其基因组变成 DNA 的形式再整合到核内基因组中。反转录病毒只能在增殖的细胞中复制，因为它进入核中需要细胞通过有丝分裂，这时病毒基因组才可接触核的物质。

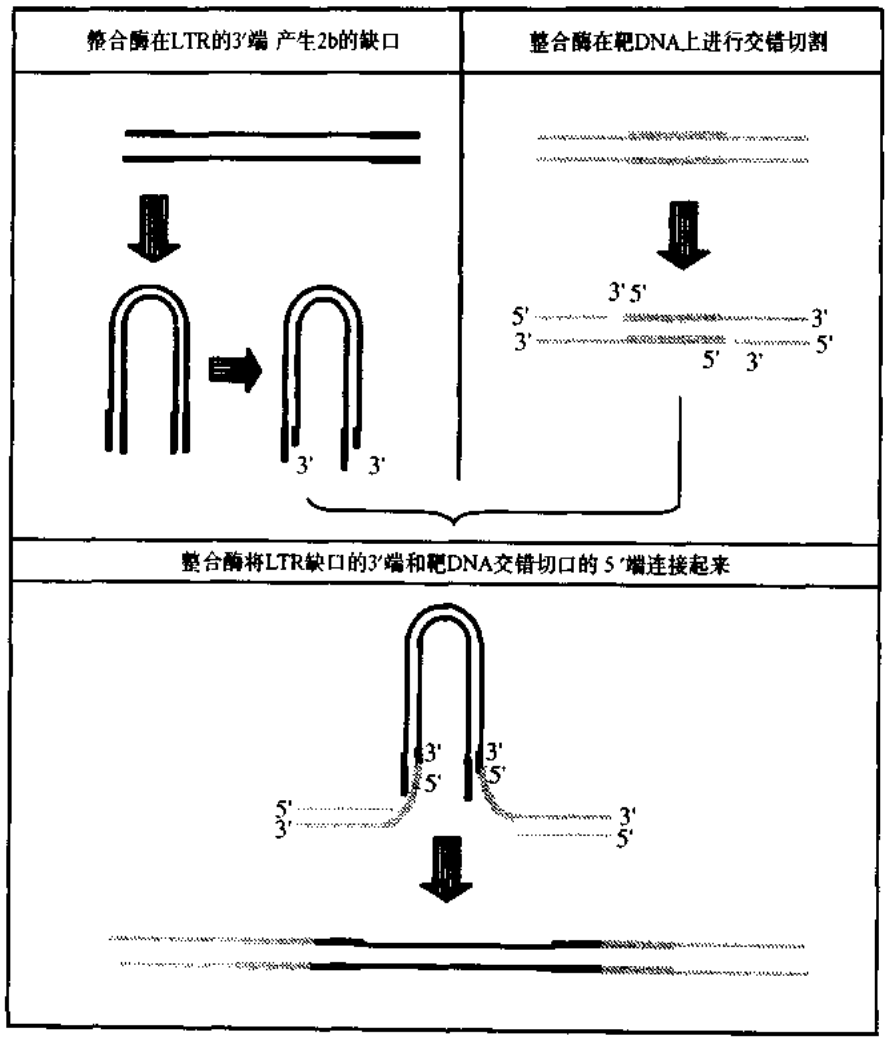


图 23 - 65　整合酶催化反转录病毒的 DNA 整合到宿主的基因组中。
（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 18 - 24）

每个 LTR 的 U3 带有一个启动子。左侧 LTR 中的启动子负责起始前病毒的转录。前病毒产生的 DNA 其 U3 序列位于左侧末端是很必要的。看来这个启动子实际上通过 RNA 反转录 DNA 时产生的。

有时（可能较为稀少的情况）右侧 LTR 中的启动子发起与整合位点相连接的宿主序列的转录。这个 LTR 也携带了一个增强子，它不仅作用细胞的序列而且也作用病毒的序列。反转录病毒的整合，以这种方式激活了某些类型的基因时可引起某些宿主的细胞发生癌变。

整合的前病毒可以从基因组中切离出来吗？ 同源重组可以发生在前病毒的 LTRs 之间。单独一个 LTR 可能作为切离的遗留存在于某些细胞的基因组中。

在反转录病毒的感染周期中，整合对于进一步产生 RNA 拷贝来说是必要的。但当病毒 DNA 在生殖细胞中整合时，它成为生物一个可遗传的“**内源前病毒**”（endogenous provirus）。内源前病毒通常不表达，但有时可被外部因素所激活，如另一些病毒的感染。

4. 反转录病毒可转导细胞的序列

在病毒的生活周期没有谈到**转导病毒**(transducing virus)的发生。而这种病毒是很有趣的。它带有从细胞中获得的序列(图 23 - 66)。病毒的部分序列被 v-*onc* 所取代。蛋白质合成时产生了一个 Gag-v-Onc 蛋白,取代了通常情况下产生的 Gag,pol 及 Env 蛋白。这个病毒是一个复制缺陷型(replication - defective);依靠它本身是不能完成细胞周期的。但在助病毒(helper virus)的帮助下,为其提供它本身已缺失的病毒的功能,它仍可不断增殖。

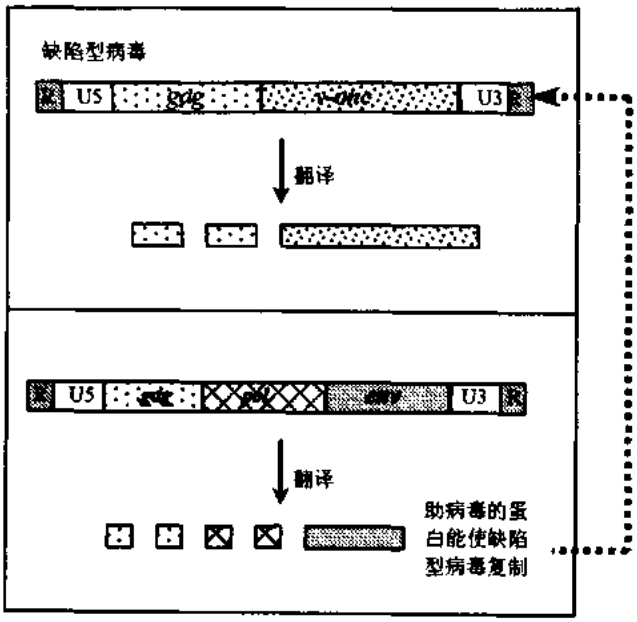


图 23 - 66　助病毒的产物可供缺陷型病毒复制。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 19 - 9)

病毒所携带的 v-*onc* 能转化一定类型的宿主细胞。前面已介绍 *onc* 来源于动物(见第 21 章),那么反转录病毒怎样获得 *onc* 的呢? c-*onc* 和 v-*onc* 的结构有着一定的同源性。c-*onc* 通常含有内含子,在细胞中低水平表达。而 v-*onc* 不含内含子,且作为病毒转录单位的一个部分高水平表达。以上表明 v-*onc* 基因来源于 c-*onc* 基因的 RNA 拷贝。

图 23 - 67 所示的是一个转化病毒形成的模型。一个反转录病毒整到 c-*onc* 基附近的位点。由于发生缺失使前病毒和 c-*onc* 基因融合在一起,转录后产生一个相连的 RNA,在其一端含有病毒的序列,而在另一端含有 c-*onc* 的序列。经剪接切除了病毒和 c-*onc* 内含子,此 RNA 具有一个信号可被包装进病毒颗粒中;如果细胞里还有另一个病毒的话产生一个完整的 RNA 拷贝。某些二倍体病毒颗粒可能含有一条融合的 RNA 和一条正常的病毒 RNA。

在二者之间的重组将会产生**转化基因组**(transforming genome),在这个基因组中病毒的重复序列位于两端。在反转录感染周期中重组以很高的频率发生,现在还不清楚这种重组是否需要同源序列。据推测可能是通过与病毒重组相同的机制进行非同源重组的。

完整反转录病毒的共同特点表明它可能从一个单个的祖先中获得。最初的 IS 因

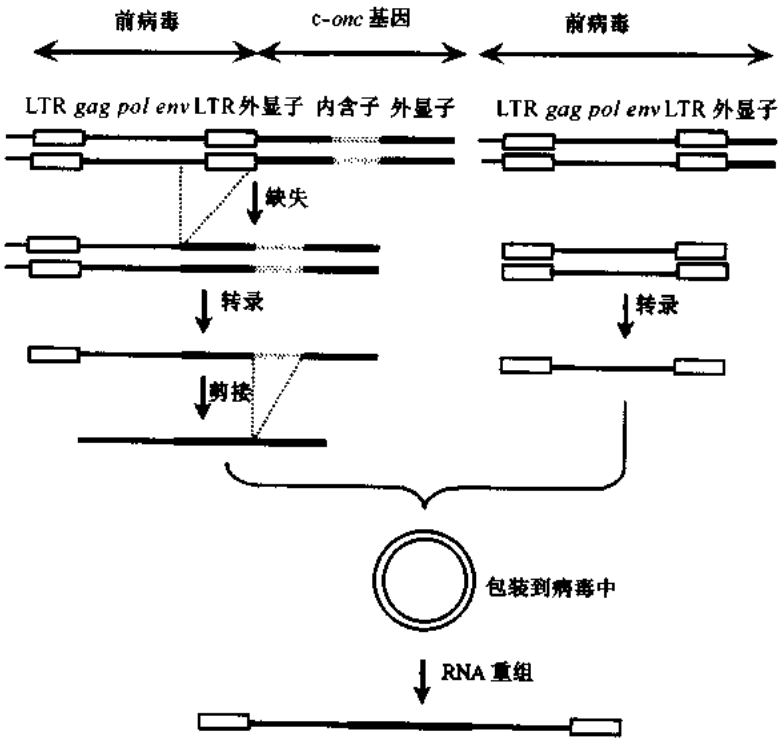


图 23－67　复制－缺陷病毒产生的途径。可能是通过一个病毒基因组的整合和产生一个病毒和细胞的融合转录本，这个转录本和一个正常的病毒 RNA 组一道包装到病毒颗粒中，经非同源重组产生了复制缺陷转化基因组。(B. Lewin: Genes V, 1995, Fig. 35－9)

子可能包围了一个宿主的核酸聚合酶基因，形成了 LTR-pol-LTR 的单位。通过获得处理 DNA 和 RNA 更为复杂的能力，还包括一些负责包装 RNA 的基因，而进化为一种感染病毒。其他的功能如转化基因可能是晚些时候含 λ 相关基因所致。病毒从细胞中获得的并不是仅是 *onc* 基因。还有其他的基因，病毒带有这些基因可能具有一种选择优势，因为它们可以刺激细胞生长。

二、 酵母的 Ty 因子

Ty 因子是由分散的 DNA 重复序列家族组成，在不同品系的酵母中它们所处的位点不同。Ty 是酵母转座子(transposon yeast)的缩写。在转座的过程中产生了一个特征性的痕迹：在已插入的 Ty 因子两侧存在 5bp 的靶 DNA 正向重复序列，Ty 的转座频率要比细菌转座子低，约为 10^{-7} ~ 10^{-8}。在单个 Ty 因子之间可能有歧化。大部分 Ty 因子可分为两类：Ty1 和 Ty917。它们的大体结构相同(图 23－68)，每个 Ty 因子长 6.3 kb，两端有 330 bp 的正向重复，叫做 δ。各种类型中的每个 Ty 因子都有一定的差异，包括酵母基因组中约有 30 个拷贝的 Ty1 和 6 个拷贝的 Ty917。另外有约 100 个拷贝的独立 δ，称为**独立的 δs**。δ 序列也显示出异质性。一个 Ty 因子的两个重复序列往往相同了或者密切相关。和 Ty 因子的连接的 δ 比单独存在的 δ 保守得多。表明对重复序列的识别可能与转座有关。

Ty 因子可被转录成 2 种带有 poly(A)$^+$的 RNA，其含量占单倍酵母细胞总 mRNA 的

5%以上。这两种 RNA 的转录都是在左边 δ 因子内的启动子中起始，一个在 5kb 之后终止；另一个在 5.7kb 长后终止，位于右端的 δ 序列内。

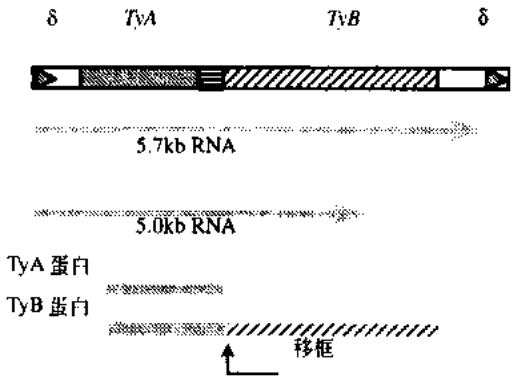


图 23－68　Ty 因子末端有短的顺向重复顺序，Ty 转录成两个重叠的 RNAs。它们含有两个阅读框，相当于反转录病毒的 *gag* 和 *pol* 基因。（参考 B. Lewin: Genes Ⅵ, 1997, Fig. 19－11）

Ty 序列有两个开放阅读框，以相同方向表达，但读框不同，有 13 氨基酸的重叠。*TyA* 的序列编码一种 DNA 结合蛋白。*TyB* 序列含有的区域与反转录病毒的反转酶、蛋白酶及整合酶序列具有同源性。

TyA 和 *TyB* 的组成与功能与反转录病毒的 *gag* 和 *pol* 相似。*TyA* 和 *TyB* 在两种阅读框中表达。TyA 蛋白由 *TyA* 读框编码，在其末端终止。但 *TyB* 读框仅表达成为连接蛋白的一部分。通过特殊的移框通读了终止密码，将 *TyA* 区域和 *TyB* 区域融合在一起。这和反转录病毒中 *gag－pol* 的翻译相似。

Ty 因子提供了一个较小的同源区。此同源区也就是由宿主系统介导的重组事件的靶子。这种重组常导致染色体的损伤。当重组发生在同一条染色体上的两个 Ty 因子之间时，会导致缺失或倒位；当重组发生在不同染色体上的 Ty 因子之间时会导致更为剧烈的改变。

Ty 因子之间的重组看来是突然发生的，当检测到一个 Ty 因子时，发现另一个的可能性就大大增加。基因的转变发生在不同座位的 Ty 因子，结果是一个因子被另一个因子所取代。

Ty 因子通过 δ 序列之间的同源重组可以切离（图 23－69）。染色体上存在大量单独的 δ 因子可能就是切离后留下的“脚印”。切离这一特点可能和 Ty 插入所引起突变的恢复有联系；回复的途径可能取决于左侧的 δ 序列。

一个独特的观点认为两个 δ 因子具有相同的序列。在 δ 中的一端存在一个有活性的启动子，在另一端存在有活性的终止子。在其他的转座因子中发现相似的特点，包括反转录病毒。

Ty 因子是典型的**反转录子**（retroposon），由 RNA 介导转座。图 23－70 显示证实此结论的实验。此实验是将内含子插入到一个 Ty 序列中。此序列由质粒上的 *gal* 启动子来控制，然后导入到酵母细胞中，转座的结果酵母的基因组中存在多个拷贝的转座子，但它们都没有内含子。我们知道只有一种途径可以切除内含子，那就是 RNA 的拼接。

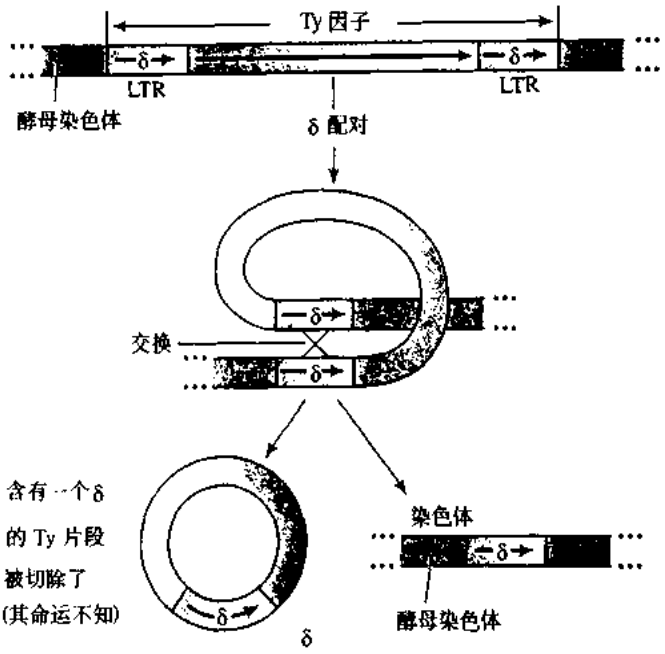


图 23－69　Ty 因子通过反向重复元件 δ 之间的同源重组而被切除，在原来的 Ty 插入位点留下了一个 δ 元件。（仿 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 20－17）

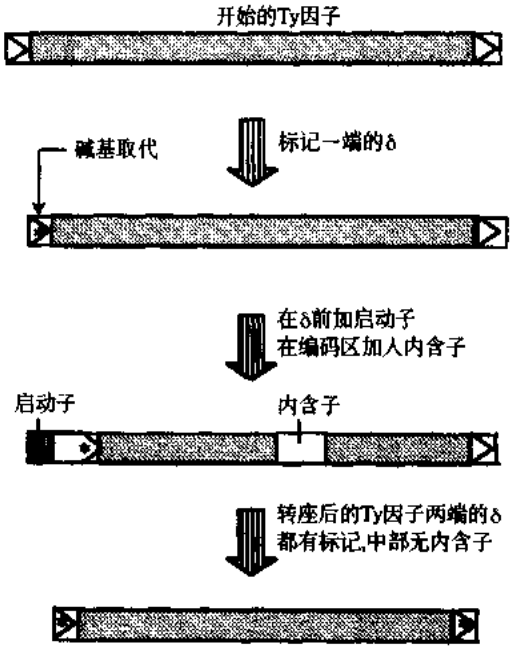


图 23－70　用基因工程的方法在 Ty 因子内加入内含子，并对左端的 δ 进行标记，转座后产生的拷贝失去了内含子，而末端的 δ 都有标记，表明来自一端。转座的频率和启动子的活性相关。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 19－12）

这表明 Ty 的转座机制和反转录病毒是相同的。Ty 因子被转录成 RNA 并被拼接装置所识别。被剪接的 RNA 被一种转录酶所识别，以其为模板合成了双链 DNA 拷贝。

Ty 因子与反转录病毒有 4 点相似：

(1) 原来的 Ty 因子两端的 δ 序列是不同的，但转座后的 Ty 两端具有相同的 δ 序列，它们来自原来的 5′端 δ 因子。如果我们将 δ 看做是一个由 U3－R－U5 组成的 LTR 的话，那么 Ty RNA 的延伸是由 R 区到 R 区。正如图 23－59～64 表示的那样，LTR 是由于在 3′加 U5，在 5′端加 U3 形成的。

(2)转座是由 Ty 因子内的基因控制的。*gal* 启动子用来控制被标记的 Ty 因子的转录，这个启动子可以被诱导，它可以通过加 Gal 来打开。启动子的诱导有两种作用：①可促进目标 Ty 因子的转座；②观察启动子的活性对 Ty 因子在酵母染色体上转座频率的影响。这就意味着 Ty 因子的产物能反式作用于其他的 Ty 因子(实际是作用于它们的 RNAs)。

(3) 虽然 Ty 因子不产生感染颗粒，但在经诱导发生转座的细胞中存在着 **Ty 病毒样颗粒**(VLPs，virus-like particles)(图 23－71)。它们含有全长的 RNA，双链 DNA，具反转录酶活性的产物以及带有整合酶活性的 *TyB* 的产物。*TyA* 的产物像 *gag* 产物一样被剪切成 VLP 的核心蛋白。这显示了 Ty 转座子和反转录病毒十分相似。不同之处在于 Ty 因子没有 *env* 基因，因此不能包装它的基因组。

(4) 仅有某些 Ty 因子在任何酵母基因组中都有活性；大部分没有转座能力，此和惰性的内源性前病毒相似。由于这些无活性的 Ty 因子保留了 δ 重复序列，这些序列在对活性 Ty 因子的产物作出反应过程中提供了转座的靶序列。

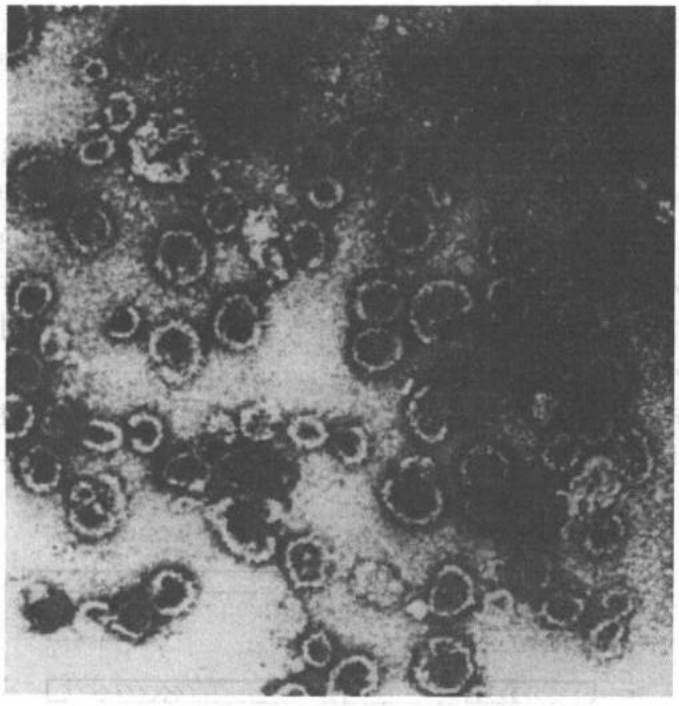


图 23－71　Ty 因子形成病毒样颗粒。(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 19－13)

三、果蝇中的反转录子

人们先观察到在果蝇中存在与 *E.coli* 插入序列相似的情况，从而推断出可能有转座因子的存在。人们发现一些不稳定的突变通过缺失，或者在原来突变位点的两端侧翼的 DNA 发生缺失而获得回复。它们可能是由各种类型转座因子所致(图 23－72)，其中包括 **copia** 反转录子，未知类型的 FB 家族和已面已介绍的 P 因子。

copia 是最为特殊的反转录的家族。它的名字反映了存在大量密切相关的编码高丰度 mRNAs 的序列。copia 家族可作为各种其他类型因子的一个范例，这些因子并不相关，但结构和行为相似。

copia 因子的多拷贝取决于果蝇的品系。一般为 20～60 拷贝。这个家族数量巨大，分布广泛。在不同品系的果蝇中它们所在的位置不同。这种差异是在长期进化中形成的。比较一下近代产生歧化的品系(超过 40 年以上)在实验室中繁殖的结果，使少数的变异得到了恢复。但其基本特点在培养的生长细胞中表露出来。每个基因组有大量的 copia 因子，大约增加了 2～3 倍。增加的因子存在于 copia 序列插入的新位点。在合适的培养条件下它们以未知的方式暂时性地增加转座频率，其范围在每基因组 10^{-3}～10^{-4}。

copia 因子长约 5kb，末端带有 276bp 的正向重复。正向重复序列本身的两端又存反向重复。在插入位点产生一个 5bp 的正向重复序列。在 copia 家族中的各成员之间的差异很小，一般小于 5%；变异体常含有小的缺失。虽然个别成员显示出很大的差异，但这些特点与其他 copia 家族都是共同的。

每个 copia 两端的正向重复序列都是相同的，这意味着在相关的过程中它们可能相互作用，或者二者都是在转座中由一祖先因子的一个正向重复序列产生的。这和 Ty 因子相似，表明了与反转录病毒的关系。

在基因组中的 copia 因子总是完整的，单个拷贝的末端重复序列未检测到(虽然我们预期它们如果产生重组的话，会缺失间插 DNA)。有时发现 copia 因子形成了环状的分子。长型的有两个末端重复，而短型的只有一个。含有 copia RNA 的颗粒倍受关注。

copia 序列含有一个长 4227 bp 的阅读框，阅读框的一部分与反转录病毒的 *gag* 和 *pol* 序列同源。值得注意的是阅读框中没有与反转录病毒的 *env* 序列同源的部分，表明 copia 和 Ty 一样不能产生病毒样颗粒。

copia 的转录本是丰度 poly(A)$^+$的 mRNA，它有两种形式，一种是完整长度的转录本，另一种是部分长度的转录本，二者有相同的 5′末端，这是由于在一个末端重复序列

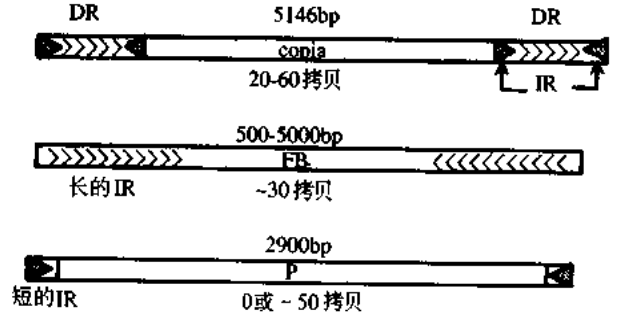


图 23－72　黑腹果蝇中的 3 种转座因子具有不同的结构。(仿 B.Lewin: Genes Ⅵ, 1997, Fig. 19－14)

的中部起始转录所致。不同蛋白的产生可能涉及到 RNA 的拼接以及多肽的剪切。

虽然我们还缺乏 copia 产生转座的直接证据。但 copia 有很多与反转录病毒结构相似之处，看来 copia 必定有一个与反转录病相关的起源。现在很难说明它具有多少多反转录病毒的功能。现在只知道它可以转座，但(和 Ty 因子的情况相似)没有证据表明它们有感染能力。

黑腹果蝇中另一个转座因子家族的成员称为 **FB**(是折回 foldback 的缩写)，它具有长度不同的反向末端重复序列。有的 FB 因子仅是由两个紧密相连的反向重复所构成。而另一些两端是反向重复序列，中间有非重复的 DNA 区域。所有 FB 家族成员虽然其反向重复序列长短各异，但都是同源的。即它的反向重复是由相同序列的 DNA 串联拷贝所组成，被不同的长度 DNA 序列所隔开。形成了 FB 的末端，相似序列单位的长度也就增加了，开始有 10bp，然后延伸到 20bp，最终达到 31bp，单个 FB 因子中的两个拷贝的反向重复序列是不同的。

关于 FB 因子的末端结构提出一个难题，现在还不知道是什么赋于它们转座的能力。有时两个不相同的 FB 因子可以协同转座一个大片段间插 DNA 序列。可能以细菌转座子相似的方式转座 (在 FB 因子之间的 DNA 的长度可能要此细菌的长得多，有的可达 200kb)。也可能是通过两个末端的 FB 因子使任何 DNA 序列折叠成一个单位进行转座。

四、两类反转录子

反转录病毒是反转录子的典型例子。因它们编码反转录酶或整合酶，所以有转座能力。反转录子不同于反转录病毒，它们自己并不形成独立的感染颗粒。但在转座机制等其他方面它们是相似的，这一组被称为**病毒超家族**(viral superfamily)。

表 23－8 表示两类不同的反转录子。

另一类反转录子根据其内外部特点，推测它们起源于 RNA 序列。我们目前只能推测它们是如何产生 DNA 拷贝的。估计转座的靶位点是通过别处编码的酶来作用的。这一类反转录子起源于细胞中转录本，并不编码具有转座功能的蛋白。这一组就称为**非病毒超家族**(nonviral superfamily)。

哺乳动物的基因组含有大量相对短的序列，它们与另一些序列相关(见第 10 章)。这些成分的有意义部分构成了反转录子。两个家族说明了大部分转座因子。开始时是一些相似的分散重复序列。每个家族的很多成员都分散在基因组中。LINS 是指长的分散序列

表 23－8 反转录子分为病毒超家族和非病毒超家族

	病毒超家族	非病毒超家族
共同类型	Ty(酵母)	SINES B1/Alu(哺乳动物)
	copia(果蝇)	pol Ⅲ转录的加工假基因
	LINES L1(哺乳动物)	
末端	长末端重复序列	无末端重复序列
靶重复序列	4～6 bp	7～21bp
阅读框	反转录酶和/或整合酶	无(不编码和转座有关的蛋白)
组成	可含内含子(在亚基因组 mRNA 中被切除)	无内含子

引自 B. Lewin: Genes Ⅵ. 1997, Table 19.1

(long interspersed sequence), SINES 是指短的分散序列(short interspersed sequence), 更为重要的区别是 LINEs 是由 pol Ⅱ转录的, 而 SINEs 是由 pol Ⅲ转录的。

哺乳动物基因组含有 20～50 000 拷贝的 LINES, 称为 L1。其典型的成员长约 65 000bp, 尾部有多聚腺苷, 中部可能存在阅读框。一种因子具有 1137 和 3900bp 的两个阅读框, 二者有 14bp 的重叠, 有转录本存在。重复 DNA 的存在意味着 LINES 家族在单个成员中有变异存在。但在一个物种中家族成员中具有一定的同源性, 这种同源性要比种间同源性高。

图 23－73 表示超病毒家族成员与反转录病毒的比较。我们知道一个有活性的 Ty 因子编码转座酶的功能。我们可以推论果蝇基因组中的 copia 序列可编码转座功能蛋白的某种活化因子。

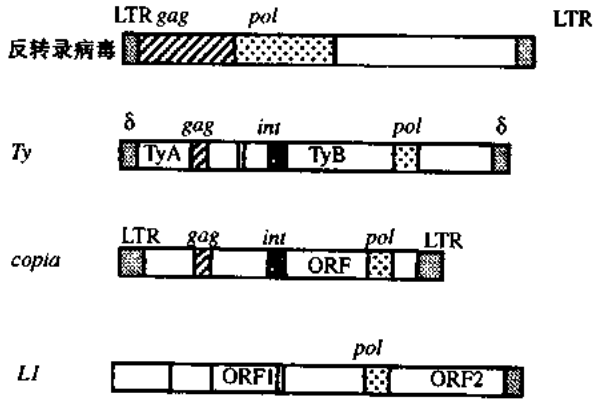


图 23－73　病毒家族的反转录子末端有重复顺序内部有开放读框。
(仿 B. Lewin: Genes Ⅵ, 1997, Fig. 19－15)

LINES 因子原来是划在非病毒超家族中的, 但测序后发现 L1 中的一个开放读框和反转录病毒同源, 这表明 L1 原来可能是一个编码其特有转座酶的可移动基因。因此将它放在病毒超家族。但还不知道 LINES 因子是否编码有功能的蛋白, 或者仅仅只是保留了与反转录子的同源性。

LINES 和一些其他的因子的两端都没有典型的反转录病所特有的 LTR。反转录的引发并不涉及 tRNA 引物与 LTR 配对的反应。这些因子中的开放读框失去了很多的反转录病毒所具有的编码序列, 如蛋白酶、整合酶。但具有典型的反转录酶样的序列, 其产物可能有内切酶活性。

图 23－74 表示无 LTR 的反转录转座子怎样进行转座。反转录转座子编码的具内切酶活性的产物在靶位点切开双链, 此因子产生的 RNA 在切口处与蛋白结合, 裂缺为在 RNA 模板上合成 cDNA 提供了 3－OH。另一条 DNA 链的切开使其打开, RNA/DNA 杂种双链或在这个阶段或在变成双链 DNA 之后与缺口的另一端连接。相似的机制也可用于某些可移动的内含子中。

由于 LINES 来源于 RNA pol Ⅱ的转录本, 其基因组序列必然是失活的。它们缺乏转录起始点上游的启动子, 因此它们一般具有成熟转录本的特点, 它们被称为**加工假基因**(processed pseudogenes)。

图 23－75 将加工假基因和原来的基因及 mRNA 进行比较以显示其特点。但各个例子中可能并不具备图中所示的所有特点。RNA pol Ⅱ的各种转录本原则上都可能产生

这样的假基因。包括被加工的珠蛋白假基因。它是最早被发现的。假基因的起始点相当于RNA的5′端，因此可以推测此DNA也许是来源于RNA。各种假基因都没有内含子，由外显子直接连接而成。我们知道不存在识别DNA中内含子的机制，因此这一特点表明可能存在在一个RNA介导的阶段。假基因的末端有一段A：T碱基对的连续序列，推测可能来自RNA poly(A)尾巴；另一方面假基因有一个短的正向重复，推测可能是经过一个转座式的行为而产生的。加工假基因所处的位置和它们原来的位置是无关的。

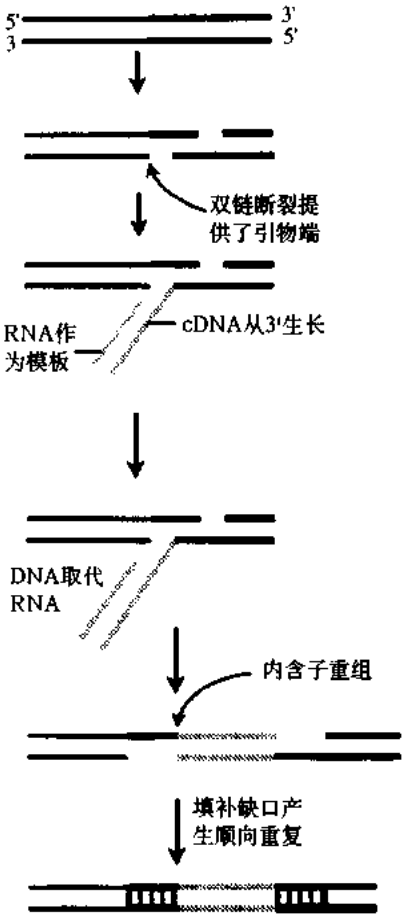


图23－74　无LTR的反转录转座子通过切开靶位点提供了在RNA模板上合成cDNA的引物。(B.Lewin: Genes Ⅵ, 1997, Fig. 19－16)

加工假基因不带有任何可用于发起转座或执行RNA反转录的遗传信息。此过程是否可能由反转录病毒来介导呢？它是否由一些异常的细胞系统来完成呢？转座序列的末端多半偶然地与转座子末端序列相似。

通常在这些基因组中会发生转座吗？或者要问我们所见到的假基因等仅仅是古代系统的"足迹"？转座能保留下来它们必然在生殖细胞中发生。相似的事件可能也会在体细胞中发生，但不会传给下一代。

最为突出的SINES家族成员的比较，它的长度最短并且高度重复，使得它可与简单序列的DNA相比，除了这个家族的个别成员之外都以基因簇的形成串联排列在基因组中。在一个物种中家族成员相似性要高于

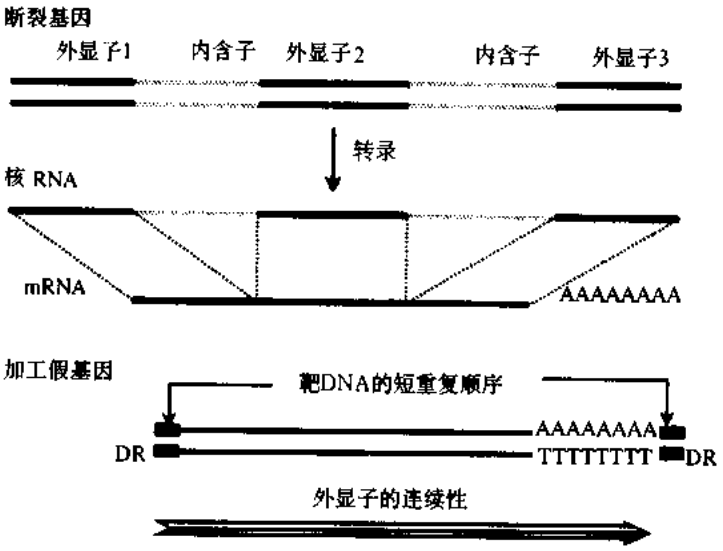


图23－75　假基因可能是由RNA反转录成DNA双链后再整合到基因组中产生。(仿B.Lewin: Genes Ⅵ, 1997, Fig. 19－17)

种间家族成员的相似性，这是很有意义的。

在人类的基因组中大部分中度重复 DNA 以 ~ 300bp 的序列存在，分散排列在非重复序列 DNA 中，其中至少有半数是可以被限制性内切酶 *Alu* Ⅰ在 170bp 处的单个位点上剪切。所有能被剪切的序列都属于一个家族的成员，称为 ***Alu* 家族**(Alu family)。在单倍体基因组中约有 30 万个成员，大约每 6kb DNA 中就有一个 Alu 序列。单个的 Alu 序列广泛地分散在基因组中。在小鼠中存在其相关的重复序列家族称为 ***B1* 家族**，约有 5 万个成员。在中国仓鼠中称为 ***Alu* 等同家族**，在其他动物中也存在 *Alu* 相关序列。

单个的 *Alu* 家族的成员是相关的，但不相同。在人类中 *Alu* 家族由 130bp 的串联重复序列组成。有 31bp 的无关序列在其二聚体的右半侧，这两个重复序列有时称为 *Alu* 序列的"左半边"和"右半边"。*Alu* 家族的成员之间序列保守有平均 87%的同源性。小鼠的 *B1* 重复单位长 130bp，相当于人类 *Alu* 序列的一个单体，与人类 *Alu* 序列有 70% ~ 80%的同源性。

Alu 序列是和 7SL RNA 相关的，7SL RNA 是蛋白转运中信息识别蛋白的一个成分(见第 15 章)它和 *Alu* 右半的单体同源。7SL RNA 的编码基因是 pol Ⅲ转录的，这些基因或与其相关的基因有可能产生失活的 *Alu* 序列。

Alu 家族的成员与两端有短的正向重复的转座子相似。但它们出现了一种奇怪的特点，那就是家族中的各成员的重复序列长度不同。由于它们来自 RNA pol Ⅲ的转录本，所以有可能个别成员中带有内部具有活性的启动子(见第 13 章)。

Alu 家族的各种特点已被发现，而且已知其广泛存在表明它们可能有某种功能，但还不清它们的作用是什么。

Alu 序列具有一个 14bp 的区域，此区域和乳头多瘤空病毒 (如 SV40) 及乙肝病毒复制区域中存在的序列相同，有这样一种可能，*Alu* 家族与真核的复制区有关系，但家族的成员数与此不符合，*Alu* 序列的数目要比复制区的数目高 10 倍。

至少 *Alu* 家族的某些成员是能被转录成独立的 RNAs。在中国仓鼠中 *Alu* 等同家族的某些成员(不是所有的)在体内存在转录本。在其他转录单位的附近发现了它们的短转录单位。

结构基因的转录单位中包含着 *Alu* 家族的成员。在单个核分子中存在的多拷贝 *Alu* 序列可能产生二级结构。其实 *Alu* 家族成员以反向重复序列的形式存在是哺乳动物核中 RNA 产生二级结构的重要因素。

五、内含子的归巢

在Ⅰ类Ⅱ类的某些内含子中含有开放读框，可产生具有三种功能的蛋白。这些蛋白可使内含子(或以其原来的 DNA 形式，或作为 RNA 的 DNA 拷贝)移动(mobile)，使内含子可插到一个新的靶位点，这个现象叫做**归巢** (homing) Ⅰ组和Ⅱ组内含子分布很广，在真核和原核细胞中都有发现。

在这些内含子的开放读框中编码具有三种功能的蛋白，这些蛋白与 DNA 或 RNA 的代谢有关。①核酸内切酶：在 DNA 的靶位点剪切，使内含子得以插入；②反转录酶：涉及将内含子 RNA 变成 DNA 拷贝；③成熟酶：从前体的 RNA 中切掉内含子的部分。

第Ⅰ组内含子具有编码核酸内切酶的开放读框，有时成熟酶的活性和此蛋白有关。第Ⅱ类内含子具有编码核酸内切酶和反转录酶式的序列，成熟酶的活性也与这些

蛋白有关。在有些情况下还具有与其他酶活性相关联的成熟酶功能的遗传信息，成熟酶主要功能是使内含子的构象稳定，这于剪接来说是很必要的。

在真菌的线粒体中普遍存在着内含子的多态性或者缺乏内含子，有种观点认为内含子来源于基因的插入。通过分析酵母线粒体的大的 rRNA 基因在杂交中的重组，另一种观点认为以上的现象是基因的丢失。

在第 I 组内含子中含有编码序列。这些内含子存在于"ω^+"的酵母品系中，但另一品系(ω^-)中缺乏此内含子，将 ω^+ 和 ω^- 进行杂交其后代通常都是 ω^+。如果我们将 ω^+ 品系看做供体，将 ω^- 品系前成受体，我们从 $\omega^+ \times \omega^-$ 杂交中会发现 ω^- 基因组中产生了内含子的新拷贝，结果使全部后代都表现为 ω^+。

在两个亲本任何一个都可发生突变而抑制以上的归巢。突变的结果使杂交后代中出现了正常分离，即 ω^+ 和 ω^- 的后代数相等。这种突变表明了这个过程的性质。在 ω^- 的品系中突变发生在紧靠着内含子将要插入的位点。在 ω^+ 品系中突变发生在内含子的开放读框中，从而阻止了蛋白质的合成。这表明在 ω^+ 品系中的内含子编码的蛋白识别 ω^- 品系的靶位点，将其切开，并将内含子的一个新拷贝插入切开的靶位点中，使其变成 ω^+ 品系，而且能稳定地遗传。

ω 内含子的产物是一种核酸内切酶，它能识别 ω^- 基因并将它作为靶切开其双链。这种核酸内切酶识别 18bp 的靶序列，此序列含有内含子插入位点。靶序列的剪切是在每条链插入位点的 3′端这一侧 2 个碱基处交错切，这样剪切位点占 4 个 bp，产生了凸出的单链末端。

这种类型的剪切是和转座子移到新位点的机制有关。双链的断裂起始了基因的转变，在转变中 ω^+ 基因序列产生了一个新拷贝取代 ω^- 基因，这个过程涉及通过复制机制进行转座，且仅在 DNA 水平上发生，或以 DNA 取代的方式，或以模板转换的方式(与重组修复相类似)来完成(图 23－76)。要注意内含子的插入中断了内切酶对这个序列的识别使其稳定。

其他含有开放读框的第 I 类内含子也可以移动。内含子永久存在的机制是相同的：内含子编码的核酸内切酶剪切内含子将要插入的特殊靶位点。插入的细节有不同的特点。例如 T4 噬菌体 *td* 内含子所编码的核酸内切酶就是在靶位点上游 24bp 处剪切，然后内含子本身插入这个位点，而不是其新产生的拷贝。

虽然内含子移动的机制是共同的，但靶位点的序列和内含子编码区域之间并无同源性。据推测内含子有一个共同的进化区域，但很明显它们有很大的歧化。靶位点是在各种靶序列中最长的，对于核酸内切酶来说也是最特异的。结果内含子永久性地插入到单个的靶位点中，而且不插入基因组的任何其他地方，这就称为**内含子归巢**(intron homing)。

含有编码核酸内切酶序列的内含子在不同细菌和低等真核生物中都有所发现。由此产生一种游离的因子，其编码的功能涉及到 RNA 的剪接或者 DNA 分子之间的移动。与此相关的观点认为内含子编码区域中密码子的用法与外显子稍有不同。

在第Ⅱ类内含子中大部分开放读框都具有一个与反转录转座子相关的区域(除核酸内切酶编码区之外)。这种类型的内含子在低等真核生物和某些细菌中都有发现。反转录酶对于内含子来说是特异的，而且和归巢有关。反转录酶以初始 mRNA 为模板合成内含子的 DNA 拷贝，通过采用与反转录病毒相似的机制使内含子插入到靶位点中。和这种情况有关的此种类型的反转录转座子与失去 LTRs 的一组反转录子是相似的。在靶位点产生缺口，形成的 3′－OH 对于引发是必需的。

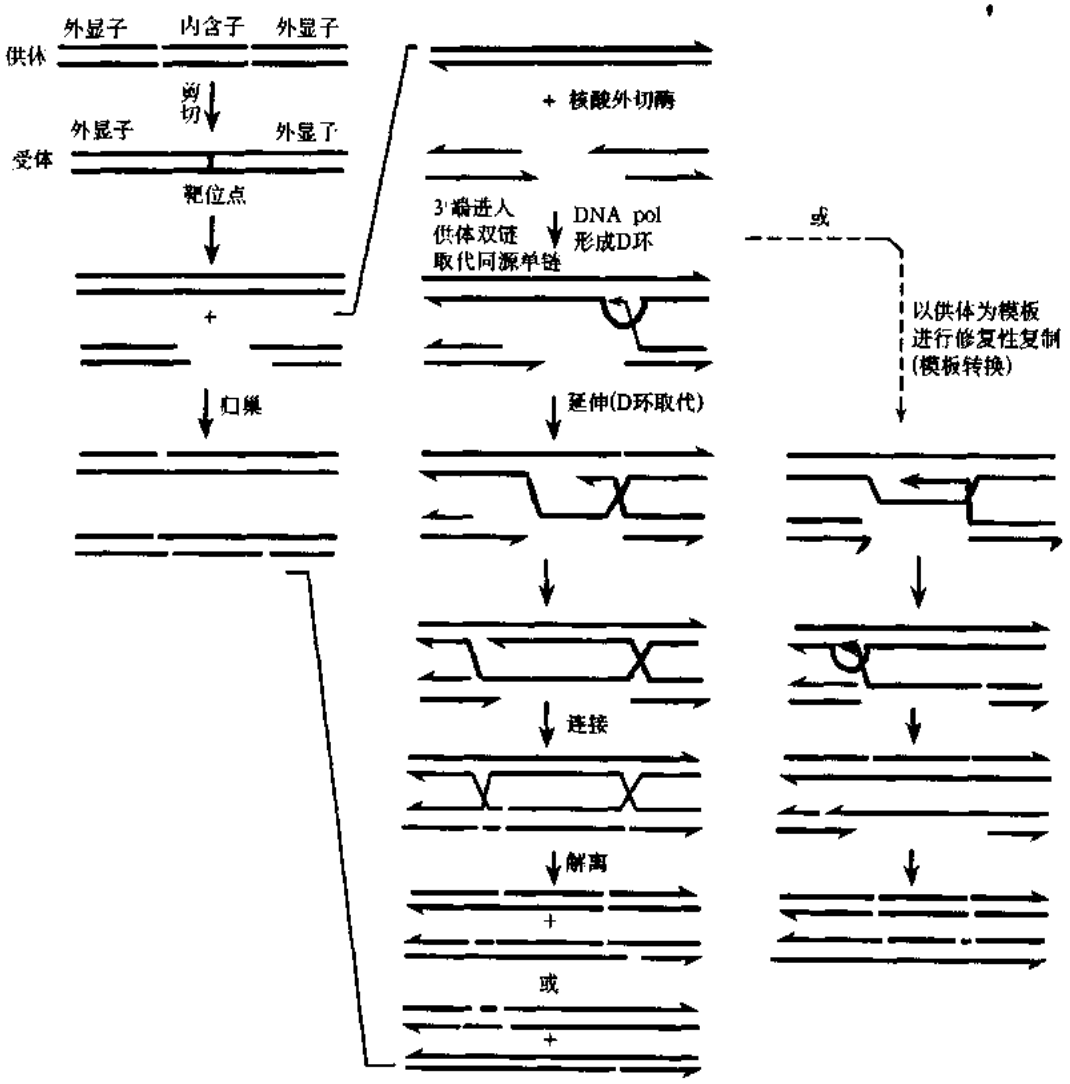


图 23－76　DNA 介导Ⅰ类内含子归巢。通过 D 环取代或模板转换来实现。
(参考 J. Bio. Chem., 1995, 270: 30237－30240)

图 23－77 表示第Ⅱ类内含子归巢的例子，与前面介绍的反转录转座子转座的机制相似，内切酶在靶位点的双链 DNA 上切一切口，在缺口上产生了 3′末端为反转录酶提供了引物。内含子 RNA 为 cDNA 的合成提供了模板。由于此 RNA 含有内含子两侧的外显子的序列，所以此 cDNA 要比内含子本身长，它能跨越双链切口，结果是使内含子插入到靶位点。

将一种核糖核蛋白和 DNA 底物一起温育在体内可产生移动系统。这个核糖核蛋白含有一个带有第Ⅱ组内含子的 RNA 和它的蛋白产物。它具有核酸内切酶活性，在恰当的靶位点交错切割双链。核糖核蛋白的 RNA 和蛋白两种成分对于剪切都是需要的。在催化中二者可能都有作用。

可移动的内含子似乎是被插入到先前存在的基因中。它们可能进化为核内前体－mRNA 内含子。具有自我剪切的能力的内含子显示了剪接进化的重要作用。例如它们能移动到核中并成为 snRNA 的祖先。

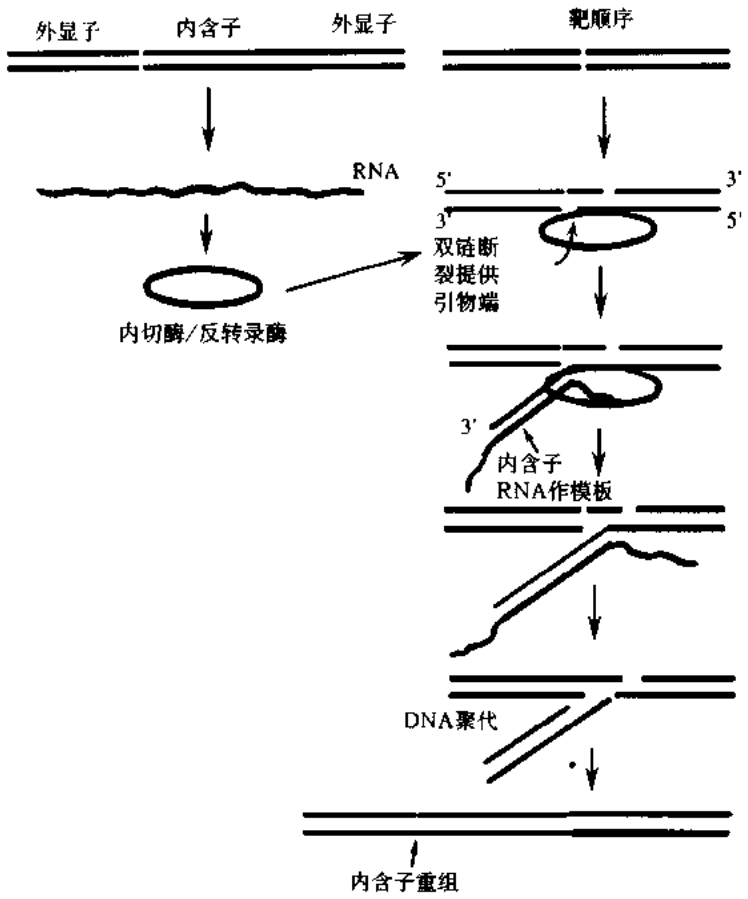


图 23－77　第Ⅱ类内含子由 RNA 介导归巢。内含子编码的反转录酶使 RNA 的拷贝插入产生双链断裂的靶位点中。（仿 B. Lewin: Genes Ⅵ, 1997, Fig. 31－11）

第二十四章

基因调控和细胞的分化发育

基本概念

无论是动物还是植物的细胞核都含有生物全部的遗传信息。细胞的分化和发育其实质是基因表达调控的差异。多利羊克隆成功说明高等动物分化的过程有的也是可逆的。

动物发育的途径是复杂的，不同的生物具有不同的模式。

C. elegans 的发育是程序化的。

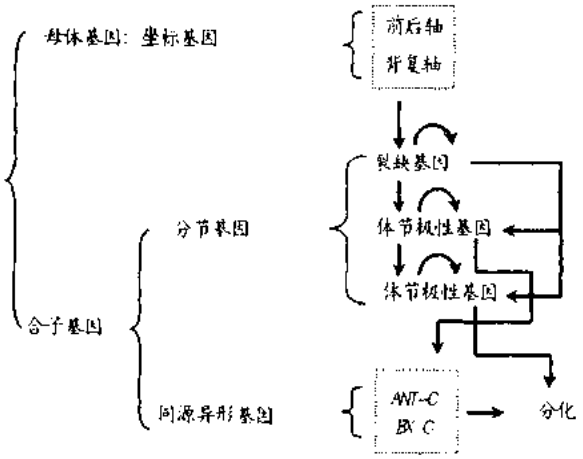
果蝇早期的胚胎发育由三类不同基因来调控：母体基因、分节基因、同源异型基因。母体基因在受精前表达，又称为坐标基因(coordinate gene)，含有 4 组：前部、后部、末端和背腹部。有的在母体的体细胞中表达，有的在母体的生殖细胞中表达，以决定胚的主要坐标轴——前后轴（前部，后部和末端）和背复轴。产物以形成素（跨膜受体和转录调节蛋白）的形式起作用。母体基因的作用是在胚胎决定定向和空间定位中的第一步，即建立前后轴梯度和背腹轴梯度。

合子基因含分节基因和同源异型基因。分节基因的主要功能是在囊胚期形成分隔（体节和副体节）及其极性。分节基因含裂缺基因、配对规则基因和体节极性基因。

同源异型复合座位主要的功能是决定每一体节形态的分化。果蝇前部体节(1～4 副体节)特征由 *ANT－C*(触角复合体)来决定，后部体节(T2－A8)的特征由 *BX－C*(双胸复合体)来决定。在 *ANT－C* 和 *BX－C* 复合座位上的异型基因簇里基因排列序列与在果蝇体内表达的位置排列序列之间具有相关性。

同源异型盒是同源异型基因中共同含有的保守序列，分节基因中有的也含有同源异型盒。同源异型盒编码的蛋白序列称之为同源异型结构域。

果蝇发育中几组基因的相互作用关系如下：



果蝇和哺乳动物的性别决定的机制不同，但都涉及到复杂的级联调控。

果蝇的性别决定主要取决于 X：A 性指数，性指数的本质反应母体效应基因，X 染色体上的几个调节基因（分子因子）和常染色上几个调控基因（分母因子）相互作用的结果。这些基因的产物多为转录因子，大部分为 HLH 蛋白，少数为锌指结构，常形成异二聚体。果蝇的性别决定中"总开关基因"是 *Sxl*，还有一组调节基因和另一些性别特异基因，形成了一系列调节级联。在调节基因中一般都是由前一个基因的产物对后一个基因的初始 RNA 进行选择性剪接，使其能成为

有效蛋白的 RNA 模板。

哺乳动物中目前除 *SRY* 以外还发现了一系列与性别决定相关的基因。这些基因大部分为调节转录因子，参与性别决定途经中的级联调节。

在所有高等生物中，生命开始于一个单个的细胞，那就是受精卵。受精卵经多次的分裂，分化形成了由各种组织系统组成的生物体。一个受精卵变成一个成体这个过程似乎十分神秘，发育生物学的目的就是要解开这个谜。

基因组编码生物体整个发育过程的潜在计划。亲代的一些基本特点可以忠实地传递给后代，这就反映了这一基本原理。如何按照潜在计划的模式形成（pattern formation）可靠地执行发育程序，构建复杂的形态呢？这个问题的解答是发育遗传学主要的目标之一，也就是在分子水平上，采用遗传学、胚胎学以及分子生物学的方法来探索这种模式形成。

在动物中有三种不同的模式形成：一种是完全程序化的。这种类型是以秀丽隐杆线虫（*C. elegans*）为代表，这种线虫从合子到具有 959 个细胞的成体整个的过程都井然有序，每个细胞的命运都早已确定，它们将在特定的时间分裂和迁移，并分化为特定的组织；第二种类型是无程序化，细胞的命运具有一定的随机性，如哺乳动物。第三种类型属于前二者之间，为半程序化。如果蝇的**成虫盘**（disk）也称为**芽盘**，此是幼虫体内含有的许多未分化的细胞小团，细胞里具有大量的细胞核和游离的核糖体，若将它分离出再移植到同龄受体幼虫中，不论移植在何处它都发育成某个特定的组织，表明它明显地受到了程序安排。

在发育过程中不同的细胞有着各自特殊的命运（fates），也就是能分化（differentiate）成特殊类型的细胞。确定这种特殊命运的作用就叫做**决定**（determination），这是个渐进的过程。对于哺乳动物来说一个细胞并不直接地从未分化的细胞（全能性，totipotent）向某个特殊方向分化。在受精卵分裂的过程中，“决定”是在每个细胞系中产生更加精确的子细胞的命运。这些“命运决定”中的有一些是由分裂细胞自己产生的；而另一些是由相邻细胞输入的一些物质来确定的。

以前讨论这些问题是从形态学的角度来描述，而现在则着重从分子水平来探讨发育和分化的实质。其中心是：①细胞分化是由什么触发；②一个个体各种细胞所含的基因组都是相同的，为什么会分化为不同类型的细胞；③分化是否可逆？这些汇集到一点就是基因表达的调控和分化、发育的关系。

第一节　高等生物的细胞核是否有全能性

一个受精卵是有全能性的：它可以产生各种类型的细胞。在发育的什么阶段细胞的全能性受到了限制呢？回答这个问题的实验方法是在各个阶段观察一个细胞或细胞核是否具有发育为一个完整个体的能力。多年来这个问题一直引起人们的关注，经多次实验的失败大部分学者认为高等生物的细胞核是没有全能性的，也就是分化是一个不可逆的过程。另一部分学者则认为高等生物细胞核应具有全能性，分化是可逆的，只是人们尚未找到适当的实验方法而已。1997 年 2 月苏格兰的 Ian Wilmnt 在 *Nantue* 杂志上发表了有关用体细胞核移植技术培育成动物——“多利”羊的文章，震动了整个的科学界和人类社会“多利”这头通过体细胞克隆产生的小羊也蜚声世界，成为最耀眼的明星。从此，上述争论的问题终于有了明确的答案。这

个答案实是来之不易。前人的多项实验的结果为今天的突破积累了很多的理论基础和经验，他们功劳也不应被遗忘。

在人类中的同卵双生子或多胞胎表明人类的遗传信息在受精卵的前几次卵裂时是精确的、忠实的。但已知早期胚胎的发育最初是受到母体基因型的控制。以后的分裂使细胞中的核产生了什么变化呢？它们仍保持着全能性吗？

很多高等生物都能产生新的器官和组织。例如海盘车丢失了“臂”或尾巴后都能再生一个。人体也能很快地修复肝脏。但这种现象只限于某些组织。在自然界中尚未发现动物的单个体细胞能产生一个完整的生物。

一、植物细胞核的全能性

植物能从一个体细胞长成一株完整的植物，但在自然界中并未发现这种现象。然而这至少表明有的植物在人类的操纵下其细胞保持着全能性。例如 1950 年 Frederick Steward 用细胞培养的方法证明了这一点。他将已分化的胡萝卜的根细胞经培养长成了一株完整的植株（图 24 - 1）。用这种方法，他通过单株植物的体细胞而获得了大量相同的植物后代，这种再生的方法现在就称为**克隆**（cloning），其实“cloning”是来自一个词，意思是植物的“扦插”（cuttings），进行营养繁殖。

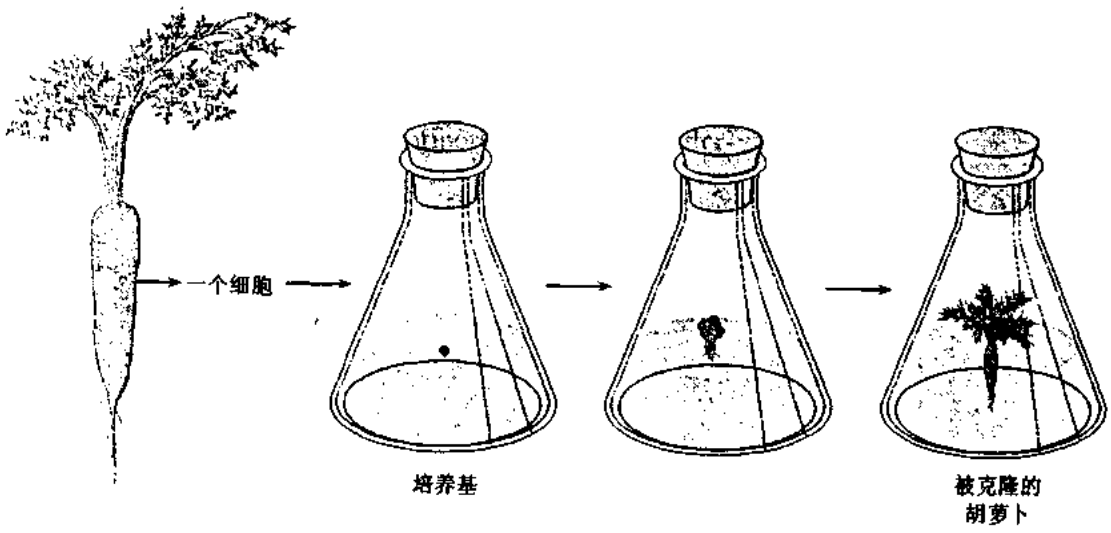


图 24 - 1 从成熟胡萝卜块根的韧皮部取一个单细胞克隆成一株完整的植株。在培养基中必须含有植物所需的几种激素。（引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5 th ed., 1993, Fig. 22 - 2）

多细胞生物的单个细胞能在培养条件下生长，而且可像微生物那样的被处理和分析。也可诱导突变，然后通过选择培养基来筛选突变细胞。这些突变细胞能经激素的作用，分化成成熟的植物。用培养的方法可简化和加速植物的繁殖。有些已分化的植物细胞可培养成植株证明了其具有全能性。那么动物细胞有相似的全能性吗？

二、两栖动物的核移植

1950 年 Robert Briggs 和 Thomas King 在研究细胞在发育中潜力的变化时建立一种

两栖动物的核移植技术。他们用一根很细的毛细玻璃吸管插到受精的蛙卵中，将细胞核吸出产生了去核蛙卵。用这种方法　就可以研究动物细胞的全能性。也便于检测一个已分化的细胞核在未受精卵的细胞质这样环境中其发育的能力。Briggs 和 King 从分化的细胞中分离核然后将其注入到无核卵中（图 24 - 2）。他们发现一个去核的卵获得到一个体细胞的核时可以分裂。而且还发现从蛙的囊胚（blastula）细胞中分离的核保持着全能性：即将从囊胚细胞中分离出的核注入到去核卵中不仅可以正常分裂，而且还能发育成蛙。

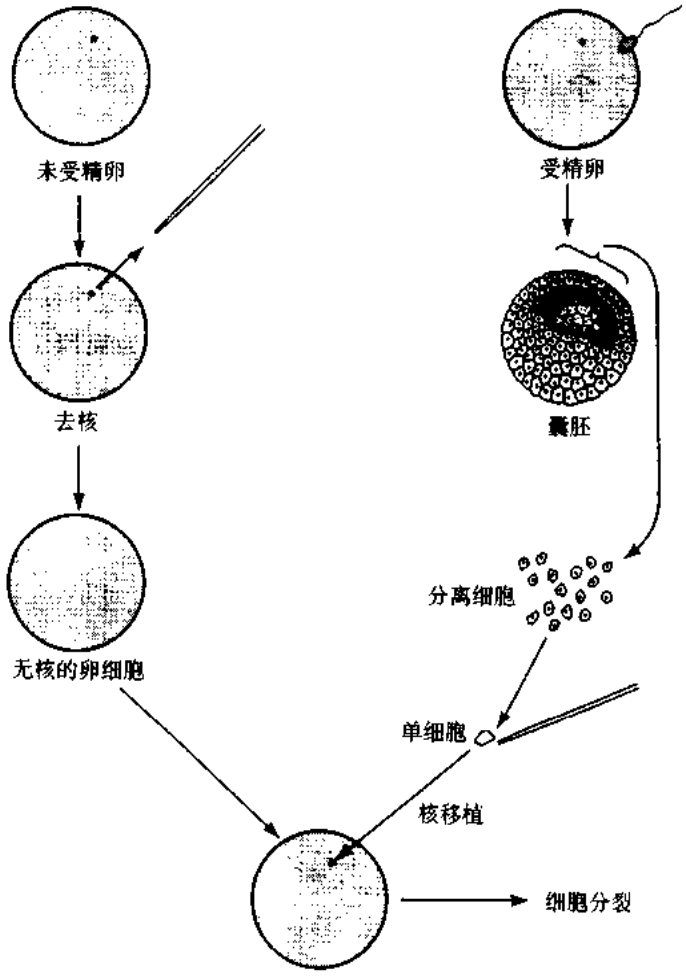


图 24 - 2　两栖动物的核移植实验。从囊胚中分离单个的细胞，取其核移植到去核的未受精的卵细胞中。（引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 22 - 3）

然而同样的实验如果是从原肠胚（gastrula）阶段（是胚胎发育的一个阶段，在囊胚期后，已初步形成三个胚层）分离核移植到去核蛙卵中结果就不能正常发育。它出现了囊胚期开始时由核决定的某些过程，而蛙体细胞核的全能性则已不复存在。

这些实验表明在囊胚期后，蛙体细胞核具有一定水平上的发育障碍。至少是在核移植实验中是这样的，但 John Gurdon 用非洲爪蟾（*Xenopus laevis*）进行实验出现了完全不同的结果。他从爪蟾高度分化的蝌蚪肠细胞中分离了核，然后注入未受精卵中，这个未受精卵本身的核已用 UV 照射使其破坏了。结果虽然很多的核移植卵未能发育或生长异常，但毕竟有约 2 % 的卵发育成正常的蛙，而且带有蝌蚪肠细胞

的遗传标记（图 24 - 3 ）。表明在脊椎动物中高度分化阶段细胞依然保持着全能性。

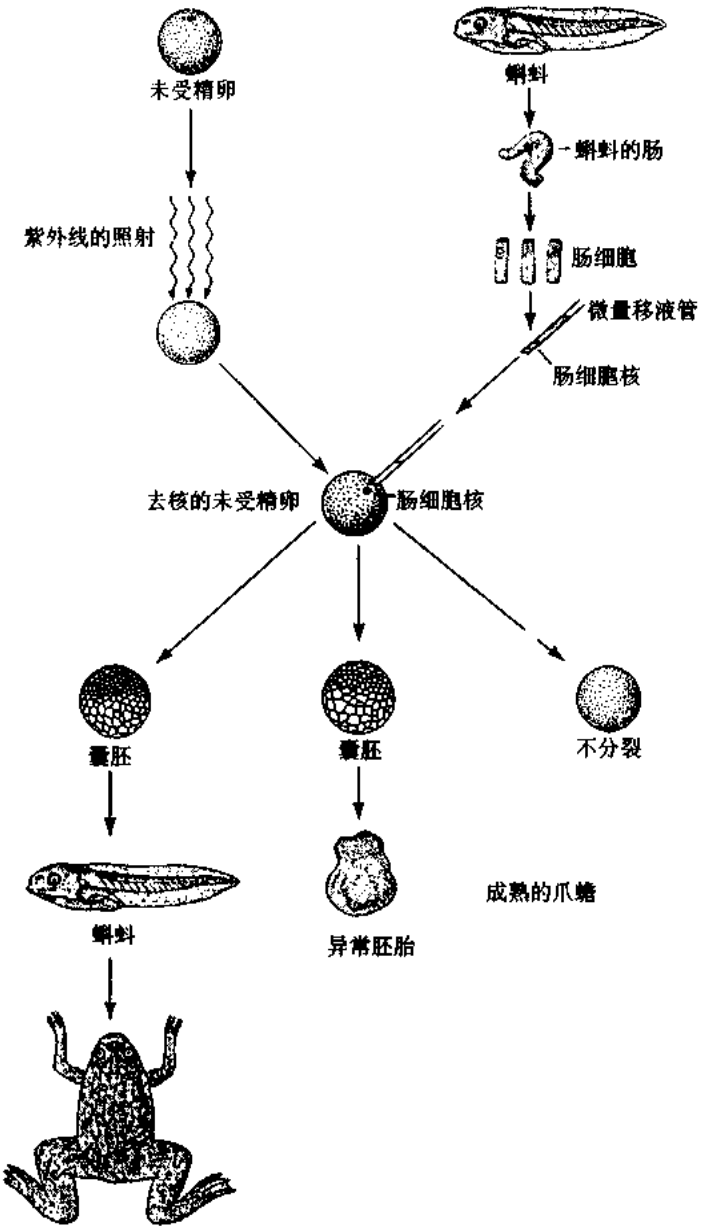


图 24 - 3 通过核移植来克隆爪蟾（Xenopus）。未受精卵通过紫外线的照射破坏其核内的 DNA(左上）。从爪蟾蝌蚪的肠中分离细胞（右），取其核移植到去核的未受精卵中。很多移核卵都能继续分裂和发育，但其中只有少数能发育成成熟的个体。（引自 Griffiths A.J.F. *et al*.: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 22 - 4 ）

三、克隆羊的诞生

1996 年 7 月一头名叫“多利”的小绵羊在苏格兰的爱丁堡诞生了，“多利”的诞生在生命科学中是又一次划时代的重大突破。

Ian Wilmut 用一头母羊乳腺的细胞核移植到另一头母羊的去核卵细胞中，经体外培养后，再移植到假孕母羊的子宫内发育而成（图 24 - 4 ）。实验一共移植了

247 个卵，多利是其中惟一成功的一头(图 24 - 5)，可见难度是很大的。

Wilmut 在进行核移植前，先将这些乳腺细胞在低血清培养液中培养(血清浓度由 100 %降低到 0.5 %)，由于血清中含有细胞周期所需的各种蛋白质因子，当浓度降低时可以抑制细胞周期的运转，使其停留在 G_0 期。而未受精的卵细胞也处于 G_0 期，这样受体和供体之间的时相是偶合的，而且处在 G_0 期时，很多基因都处于关闭状态，这就有利于细胞核移植入卵细胞质后，重新启动基因的表达，即**再程序化**(reprogramming)，从而驱动整个发育过程。这是本实验的要点之一。

另外绵羊的受精卵在第四次卵裂前核基因仍不表达，细胞分裂等活动主要是依赖于卵细胞质中储存的由母体产生的 mRNA，移入的核不需立即发挥作用，这样给核基因的表达留有了充分的时间。以便产生有关的转录因子和调控蛋白，进入核中启动和调节核基因的表达。而小鼠和大鼠受精卵中核基因打开较早，在第一、二次分裂就开始启动，这样就要求移入的核立即进入角色，因而难度较大，至今尚无一例成功。

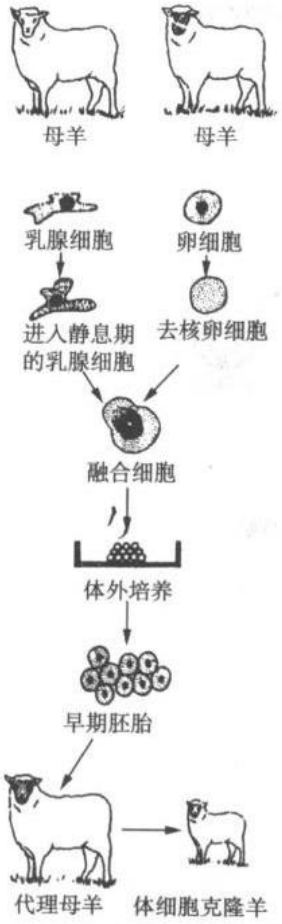


图 24 - 4　羊体细胞克隆的过程。
(引自丁小燕，自然杂志，1997，99(5)p53。)

图 24 - 5　多利羊和它的代理母亲。

克隆羊的成功有着极重要的意义：①在理论上充分证明了动物的细胞核有着全能性，发育是可逆的，从而结束了几十年来的争论；②建立了高等哺乳动物体细胞克隆的方法，为抢救濒临灭绝的珍稀动物和大量繁殖优良品种奠定了基础；③引发了人们对克隆人这一敏感问题的广泛注意。

试管婴儿和基因工程技术建立伊始都遭到了很多人的反对，但经过了 20 多年的实践人们终于改变了观念，基因工程已成为生命科学的研究、基因治疗以及生物产品的开发利用的重要方法之一。

随着科学技术的发展，克隆人的技术迟早都会建立的，这是自然科学发展的必

然。这项技术将会在器官移植等领域中造福于人类。

第二节 线虫的发育模式

遗传学和分子生物学是解决生物学问题的有力工具，对于发育生物学的问题也不例外。在过去的二十年中遗传学和分子生物学的渗入，使发育过程中很多的问题得到了解决,从而建立了动物发育分子基础的一般轮廓。

和其他生物学现象一样，某些系统对于研究发育遗传学更为有利。通过对果蝇和线虫（ *C. elegans* ）两种无脊椎动物的研究提示了部分动物发育的模式，搞清了其他一些生物中相似的发育问题。人们在思考从模型动物中获得的信息如何应用于其他系统。

发育生物学面临的主要挑战是要破译动物形态建成的分子密码，此涉及胚胎发育的早期阶段。发育生物学已从线虫和黑腹果蝇的研究中获得了一定的信息。

C. elegans 的发育

C. elegans 是一种很小的蠕虫，人们将它作为一种模式动物，其成体长仅 1mm（图 24－6)。整个的生命周期仅 3 天。野生型线虫胚胎发育中细胞分裂和细胞系的形成具有高度的程序性，这样就便于对其发育进行遗传学分析。一个成体的蠕虫仅由 959 个细胞组成。由一个受精卵发育成为成熟的成体只要两天多一点(25℃时需 52 小时)(图 24－7)。从卵到成体每个细胞的命运以及它们沿着一定的程序在特定时间的分裂和迁移都已搞得十分清楚。

C. elegans 有很多有趣的特点，它是一个染色体数很少的二倍体， $2n = 12$(有一对性染色体和 5 对常染色体），其基因组也很小，仅有 8×10^7bp，约为人类基因组的 3 %，约有 13，500 个基因。在真核生物中基因都是产生单顺反子 mRNA，惟有 *C.elegans* 与原核相似，有 25 %左右的基因产生多顺反子 mRNA(polycistronic mRNA)，这和它们通过反式剪接使下游基因得到表达有关（见第 13 章）。还有一个特点是其

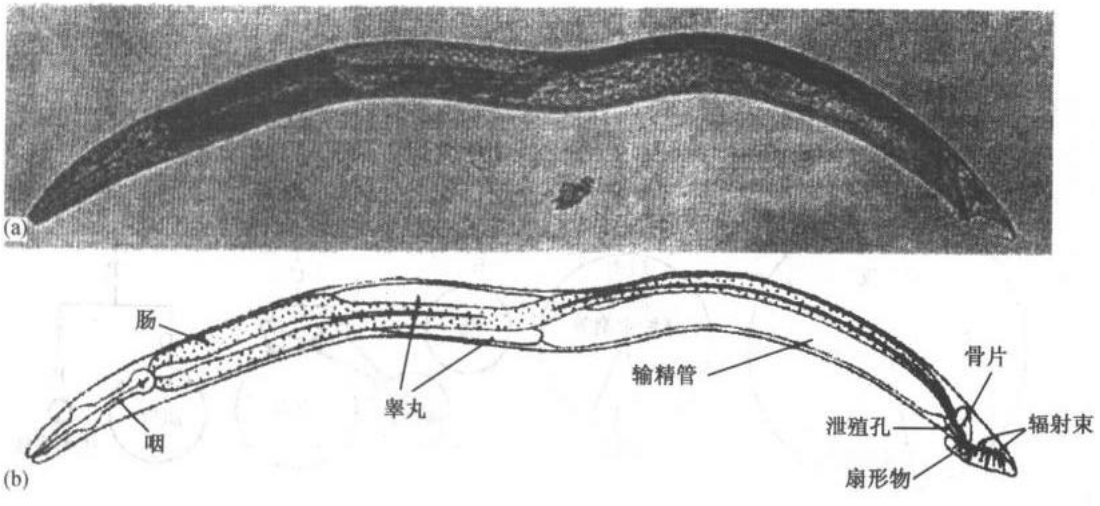


图 24－6 雄性线虫(*C. elegans*)的照片和画图。

基因组中非重复序列很高，达到 83 %，而高等的真核生物都在 50 %以下，*E.coli* 为 100 %，看来 *C.elegans* 在这些特点上都较接近原核生物，这也反映其在进化中的地位较为原始。

这种蠕虫大部分是 XX 型，是可以自体受精的两性体（ hermaphrodites ）大约每 500 个蠕虫有 1 个是 XO 型的雄体，是染色体不分离的结果。就像植物中的自花受粉一样，若有突变发生，两性体自体受精易使突变基因纯合。而互交至少要到第三代才可能出现隐性突变基因的纯合体。

C.elegans 的生活周期包括一个在卵中的胚胎发育阶段，4 个幼虫阶段和一个成虫阶段（ 图 24 - 7 ）。在受精后，前五次分裂中已建立了前后轴和背腹轴。胚胎中分裂产生的 6 个细胞（ AB 、 MS 、 E 、 C 、 D 和 P4 ）已为幼虫中的 6 个主要的细胞系统提供了前体（ 图 24 - 8 ），此细胞分裂的模式对于任何一条蠕虫来说都是一样。由于细胞分裂的“ 计划 ”决定了各种细胞在胚轴中的位置。所以细胞的位置和移动都是一成不变的。

作为一个多细胞的胚是具有三维形态结构的。是什么决定三维结构中每一个细胞的命运呢？可能是细胞的位置决定了这种拓扑学的蓝图，或者是细胞内部存在着一种对相邻细胞作出反应的能力。

在 *C. elegans* 从受精卵到幼虫、成虫，完整细胞系的建立通过细胞按预定的模式分裂和迁移才得以实现的。若从第一次分裂开始，杀死某些细胞，存活下来的其他细胞不能完全取代它们。它们发育的模式是不变的。因为分裂的过程是按“计划”进行，因此细胞的命运也是事先就已确定好了的。

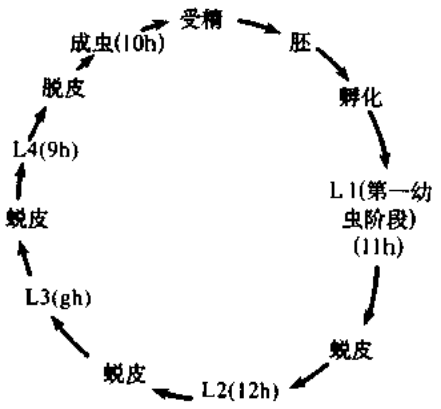


图 24 - 7　*C. elegan* 的生活周期。

L1 幼虫含有 558 个核，当发育成具两性的成虫时具有 959 个体细胞核（ 这儿提到核是因为有的细胞是“ 合胞体 ”（ syncytial ）或多核细胞 ）将要形成成体结构的胚胎细胞叫做**母细胞**（ blast cells ）（ 图 24 - 9 ），可以通过细胞系树来描述它们产生的后裔。E 细胞

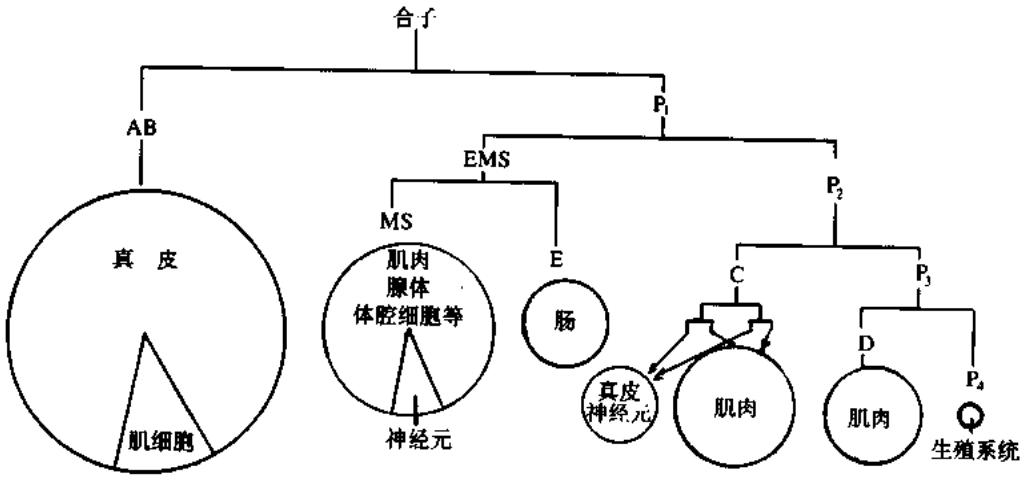


图 24 - 8　*C.elegans* 主要胚细胞系的来源。

（ 引自 Hirsch *et al*, Cold Spring Harbor Sym posium on Quantitative Biology 50, 1985, 1970 ）

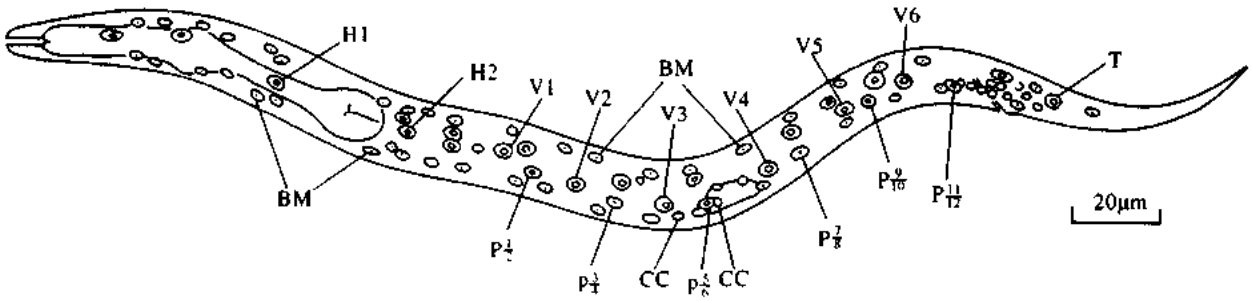
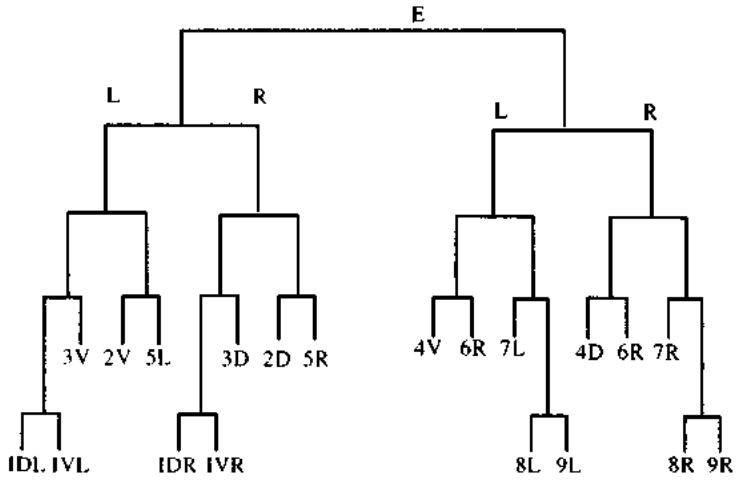


图 24－9 L1 幼虫中各种母细胞的位置。每种母细胞是一种特定组织的前体，在成体中分化成不同的细胞克隆。
（引自 Russell, P. J.: Genetics, 3rd ed., 1992, fig. 20－22）

系在胚期产生 20 个细胞以后分化为肠，其后裔从咽到尾列成一排。从细胞树我们可以看出每个细胞的空间位置，以及它们的来龙去脉(图 24－10)。



D:(dorsal)背；V:(ventral)腹；L:左；R:右
图 24－10 E 细胞系在胚期达 20 个细胞时的时空排裂。

C. elegans 存在着**时序调控基因**(chronogenes)，*Lin－14* 就是时序调控基因。*Lin－14*(o)突变型是隐性突变，突变体多蜕一次皮，使其幼虫期产生成熟表皮；另一种突变型(*lin－14*(d))是显性突变，突变体少蜕一次皮，使成虫带有幼虫的表皮。这种决定基因表达的时空模式的基因称**异时基因**(heterochronic)。它们和同源异型基因(homeotic genes)相类似。新种的进化往往来自这些基因的突变。

Lin－14 基因受到 *lin4* 基因的调节，这种调节以反义 RNA 的形式来调节的。*lin4* 编码 2 个小分子 RNA，仅有 22 和 61nt，它们的末端位点不同，它们可以和 *lin－14* RNA 3′端 7 次重复的 10nt 的序列互补，关闭了 *lin14* 的表达，机理尚不清楚(见第 16、18 章)。

第三节　果蝇的早期发育

有关果蝇早期发育的一些信息来源于经典的胚胎研究。果蝇易于培养，生活周期

短，基因组也比较小（1.4×10^8bp/每个单倍体基因组），只有 4 对染色体，且存在多线染色体，有关果蝇的遗传学知识已积累了 80 多年，遗传背景十分清楚，这些都是其重要的优势，故人们选择果蝇作为研究发育的模式动物。果蝇的卵为幼虫提供了各种营养和信息，幼虫的分化和发育也是高度精确的。

一、卵子的发生

果蝇的卵巢是由很多输卵管组成的。输卵管中的一串卵母细胞按照顺序发育。属于体细胞的滤泡细胞包被着卵母细胞和多个**滋养细胞**，滋养细胞和卵母细胞是生殖细胞（图 24 - 12）。一个单个的“细胞”经 4 次有丝分裂产生了 16 个互连细胞（interconnected cell），连接处称为**胞质桥**（cytoplasmic bridges）或称为**环沟**（ring canals）。每个细胞有 2、3 或 4 个胞质桥。具有 4 个胞质桥的 2 个细胞中的一个经减数分裂变成卵母细胞，另外 15 个细胞成为滋养细胞。滋养细胞胞质中的蛋白和 RNA 可以通过胞质桥输入卵母细胞。这种物质的积累占了卵的大部分体积，胞质连接一端接卵母细胞，这端就成了卵的前端。

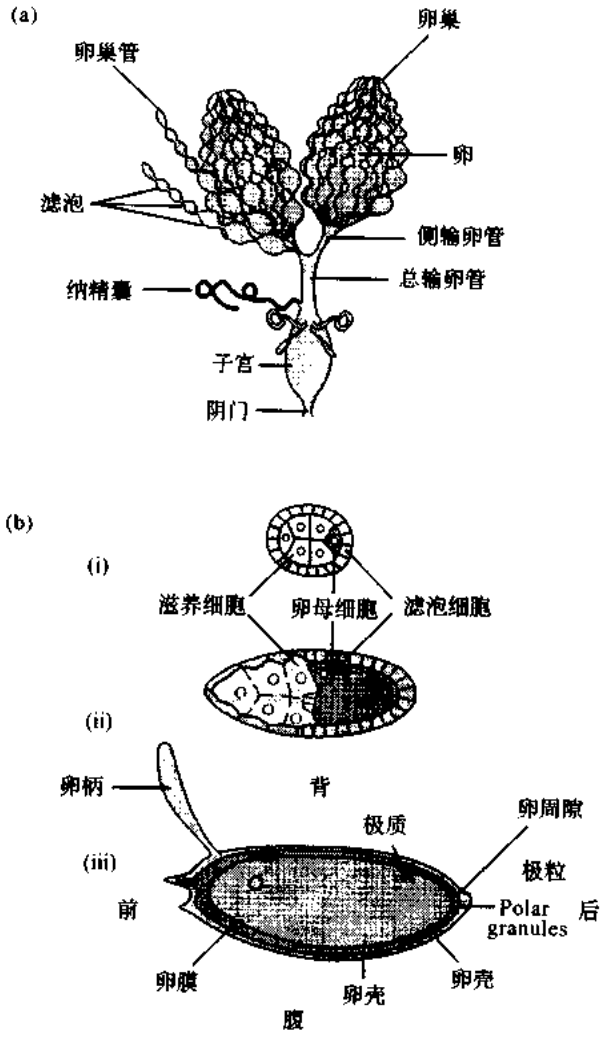


图 24 - 11　果蝇卵巢和卵母细胞的发育。（a）卵巢的结构；（b）卵母细胞的结构。
（引自 Griffiths A.J.F. *et al*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 22 - 10）

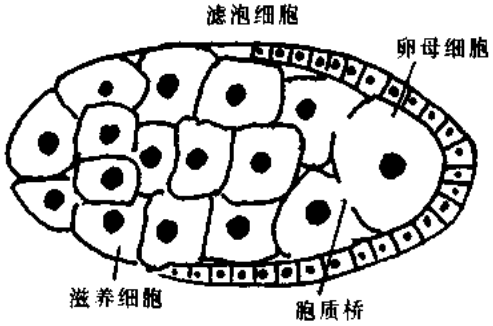


图 24 - 12 果蝇滤泡细的外表面有一层滤泡细胞，里面含有滋养细胞和卵母细胞，滋养细胞通过胞质桥彼此相互联接，也和卵母细胞相通，滤泡细胞是体细胞，滋养细胞和卵母细胞是生殖细胞。（仿 B.Lewin:Genes VI,1997,Fig.30 - 3）

二、果蝇的早期发育

受精卵开始发育产生了发育命运不同的细胞。早期发育的问题是要了解不对称性（ asymmetry ）是如何产生的。从几个细胞怎样起始分裂，从一个细胞怎样产生另一些不同性质的细胞？不同的生物不对称性也是不同的，在哺乳动物中，卵本身是同源的，产生不对称性是取决于起始分裂周期的过程。而在果蝇中卵一开始其细胞质内的成分分布就是不均一的。经过发育产生了进一步差异。

发育的分子基础将要通过基因表达的模式，即通过特殊的基因表达调控来描叙每种类型细胞的分化。控制基因表达调控的基本手段是转录，而可以调节转录的成分提供了一类重要的发育调节物。我们可以把转录调节蛋白中的活性的变化包括在内，它们可能起到改变启动子区域，调节增强子的活性，或者可能抑制转录因子的活性的作用。转录的调节蛋白常是一种 DNA - 结合蛋白，常激活特殊启动子或增强子。

这些调节物以系统的方式被打开或被关闭，形成了黑腹果蝇中决定身体各部分发育的途径。这一基本原理是卵开始时的不均一性的结果。不均一的物质成为基因控制物，卵的特殊区域获得不同的肽，这就意味着不对称的物质成为 4 个不同系统中基因表达的控制物。这 4 个系统在昆虫的卵中已被描述，它可能涉及卵中转录或翻译控制因子的定位，或涉及这些因子活性控制的定位，但结果都是相同的 - 实现基因表达的**时序调节**。

发育的起始阶段与下一阶段相连接。在起始阶段中已发生了部分胚胎决定，胚的不同区域已限定将发育成身体的特殊部分。调节此过程的基因若发生突变会导致身体某一部分的缺失、重复或发育成身体的另一部分的表型而被鉴别出来，这些基因座位成为具有“开关”功能的调节基因的主要候选者，很多已被分离，它们的产物是转录的调节物。这些基因以级联调控的方式作用另一些基因。但它们也作用另一些控制模式形成的基因。它们的靶基因常编码激酶、细胞骨架因子、分泌蛋白和跨膜受体。

观察整个的过程我们将看到在不同的胚胎区域分别建立了不同的转录模式，产生了级联调控。一个基因在某一个阶段打开或关闭，其本身又控制了另一个基因在

下一阶段的表达。其共同的特点是调节蛋白是转录因子，它们调节另一些转录因子的表达。这在果蝇和线虫的性别决定中表现得十分典型。正如在原核生物中那样调节蛋白和靶基因之间的关系是调节物识别靶基因启动子(或增强子)DNA 上的一个短的序列。对于特殊调节的靶序列都可以通过它们的一致顺序加以鉴别。

从一个受精卵发育成一个成体生物是按照一种预先决定的途径，在此途经中特殊的基因被打开，又在特殊的时间被关闭。从机制的剖析中，人们掌握了控制转录的大部分信息。当然基因调节还涉及到另一类型的信号，包括细胞群之间的细胞与细胞的相互作用。

细胞发育的机制在不同物种中是不同的，但可以用果蝇作为模型来推论发育的基本原理，以期帮助人们了解其他生物发育的机制，即级联调节决定了胚胎细胞中基因表达的模式直至发育成为成体。另外在不同种的相关生物中同源基因在发育中发挥的作用也是相关的。虽然果蝇和哺乳动物的胚胎结构是不同的，但调控的方式相同。

和发育调控有关的基因是通过突变加以鉴别的。有的突变导致在早期发育中致死，有的导致产生异常的结构。影响到身体特殊部位发育的突变引起人们的注意，因为身体某部分的结构是复杂的，需要包括很多基因在内的一套特殊基因的表达。一个基因的改变引起了整个身体某部分的结构的变化，因而可以推测这个基因可能是一个调控基因，担负着发育途经之间的“开关”或选择的作用。在果蝇中体节是作为身体的一部分来进行分析的，在成蝇中能找到它的相应部分。突变可以分为三组，通过它们对分节结构的作用来划分：①**母体基因**(maternal genes)：通过母体在卵母细胞中表达，它们可能对成熟的卵母细胞起作用或者在早期胚中起作用；②**分节基因**(segmentation genes)：在受精后表达，这些基因的突变会改变体节的数目或极性。分节基因同分为 3 组，连续作用胚胎的较小区域；③**同源异型基因**(homeotic genes)：控制体节的特征，而对其数目，大小和极性不起作用。这些基因的突变会导致身体的一部分发育成另一部分的表型。

每一组基因的作用是使胚胎特异部分的特点逐步连续地显现出来。母体基因作用卵中的广泛区域，其产物的分布差异控制了分节基因的表达；分节基因分划体节界线，同时同源异型基因控制体节的特征的发育。

果蝇的发育主要分三个阶段：卵、幼虫和成体果蝇。在发育的开始在卵中沿着前后轴(anterior-posterior)和背腹轴(dorsal-ventral)建立一个梯度。卵的前端将成为成蝇的头，而末端将成其尾。背侧在上方腹侧在下方。受精卵中的蛋白和 RNA 分子分布不均产生了这种梯度。一旦受精，此梯度立即建立起来，它控制着前后轴的发育，稍晚一些才建立背腹轴梯度。前后轴系统沿着幼虫的体长控制**位置信息**，而同时腹背系统调节组织的分化，即特殊胚胎组织的特化，包括中胚层、经神外胚层和背部外胚层。

昆虫的发育涉及到幼虫和成虫两种不同的结构。第一部分的发育是和幼虫有关。然后幼虫变态成为果蝇。这就意味着胚(幼虫)的结构和成体(果蝇)的结构明显不同。此和哺乳动物不同，哺乳动物胚胎的各个部分发育为成体的相应部分。当果蝇幼虫发育时形成幼虫身体的某一部分，但并不产生成体的组织，而身体的另一些部分将会变态成为成体的结构。虽然昆虫和脊椎动物的两者的发育是不同的，但在发育过程中存在着普遍的原理。我们就是要揭示二者发育调节的相关性。

胚胎中不连续的区域和成体身体的某一部分是相对应的。在幼虫的体表出现一些细的“节”。这是一种特殊的模式。这种模式具有前后轴和背腹轴决定的一些特

点:①沿着前后轴的"节"形成不连续的带,每一条带都和成蝇的体节相对应(图 24 - 13),即11 条带与成体的 11 个体节相对应。②沿着背腹轴,从腹部向背部延伸,腹部比较宽,背部比较细。

虽然这种带的信息只存在于幼虫的体表,但它的结构是胚中两个轴形成的全部特征。果蝇发育中很多突变表型的分析都是根据延着两个轴的带形的变异来命名。

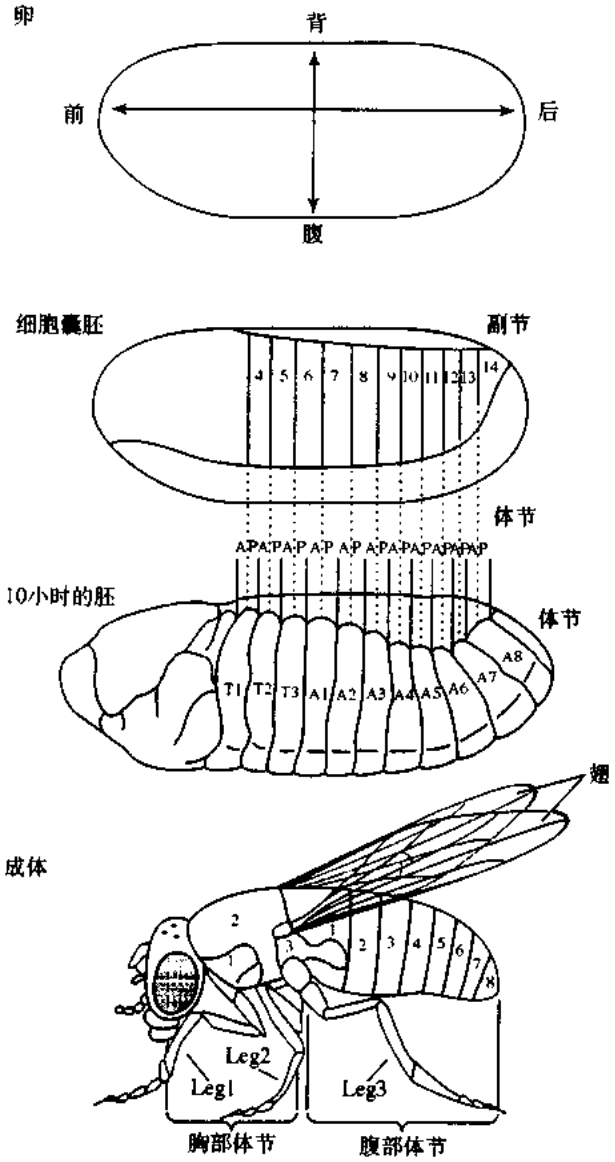
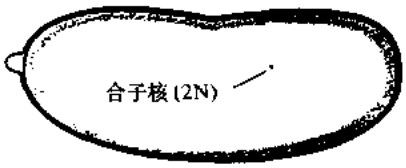


图 24 - 13 从受精卵发育成囊胚、分节胚和成体。成体的体节和胚胎的体节直接相关。
(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 17 - 28)

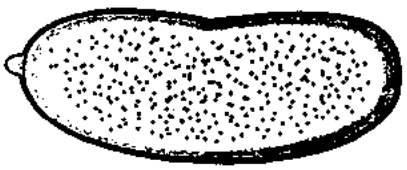
卵的梯度和成体体节之间的差异给我们提出了一些重要的问题?卵内的梯度怎样会使细胞产生不连续的差异?从一种梯度怎么会产生大量分开的分隔。梯度的特点及其影响胚胎中各种细胞发育的能力取决于果蝇发育的某些特异反应的特点。早期阶段的发育如图 24 - 14 所示:受精卵具有雄前核和雌前核两个核,后端的区域称为极质(polar plasm),前 9 次分裂核都在共同的胞质中。在第 7 次分裂时有的核就移到极质中,第 9 次分裂后核向四周迁移并分裂,在卵的周边形成了一层核,然后进行 4 次分

(a)含有两个亲体核的受精卵

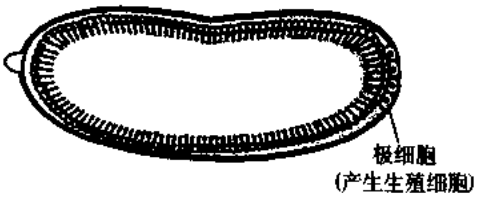
(b)亲体核融合，产生二倍体合子核



(c)核在共同的胞质中进行9次分裂，产生多核的合胞体



(d)合胞囊胚的核移向四周并分裂，延卵膜排成一层



(e)细胞囊胚。核分裂四次，核外形成膜，发育成体细胞

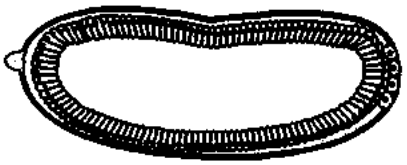


图 24 - 14　果蝇合胞体的形成和发育。（ a ）受精卵：亲体核融合产生二倍体合子的核；(c)核在一个共同的胞质中经 9 次分裂，产生多核的合胞体；（ d ）卵移向四周并分裂，沿卵膜排成一层；（ e ）核分裂四次，核外形成膜，发育成细胞囊胚的体细胞。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 17 - 27 ）

裂，以后形成体细胞。直到细胞形成前许多核仍存在一个共同的胞质中。在细胞的囊胚的阶段首先是不连续的间隔变得明显，此时卵的特殊区域被决定发育成特殊类型的特殊结构。在这个过程的一开始核迁移到卵的四周，形成了单层囊胚。但它们并不存在预先限定区域的方式。它们在囊胚的所在位置决定了它们的后代细胞将分化成什么类

型的细胞。一个核通过前后轴和背复轴的影响决定在胚中的位置和相应的行为。

第四节 母体基因的调节作用

在母体细胞中表达的基因对于早期发育十分重要，这些基因是通过雌性不育突变而被鉴别出来的。它们并不影响母体的本身，但要产生后代就需要它们。带有这种突变的雌蝇所产生的卵不能发育为成蝇；其胚能通过缺陷的表皮模式而被识别并在早期发育中死亡。

所有母体基因的共同特点是在受精前表达(虽然它们的产物可以在表达时起作用或储存起来以后使用)。母体基因根据它们表达的位点可分为两个基因家族：①在母本体细胞中表达，但影响卵发育的称为**母体体细胞基因**(maternal somatic genes)，例如它们可以作用滤泡细胞；②在母体生殖系统中表达的基因称为**母体生殖细胞基因**(maternal germ line genes)，这些基因可以作用滋养细胞，或可作用卵母细胞。有的在两个方面都起作用。

胚胎特殊区域的发育与 4 组基因有关，通过母体基因的突变可鉴别出这些基因。每一组基因可按它们的序列组成一种途径。通过常规的遗传学测验(如比较双重突变和单个突变的特点)或者通过生化的方法来检测。

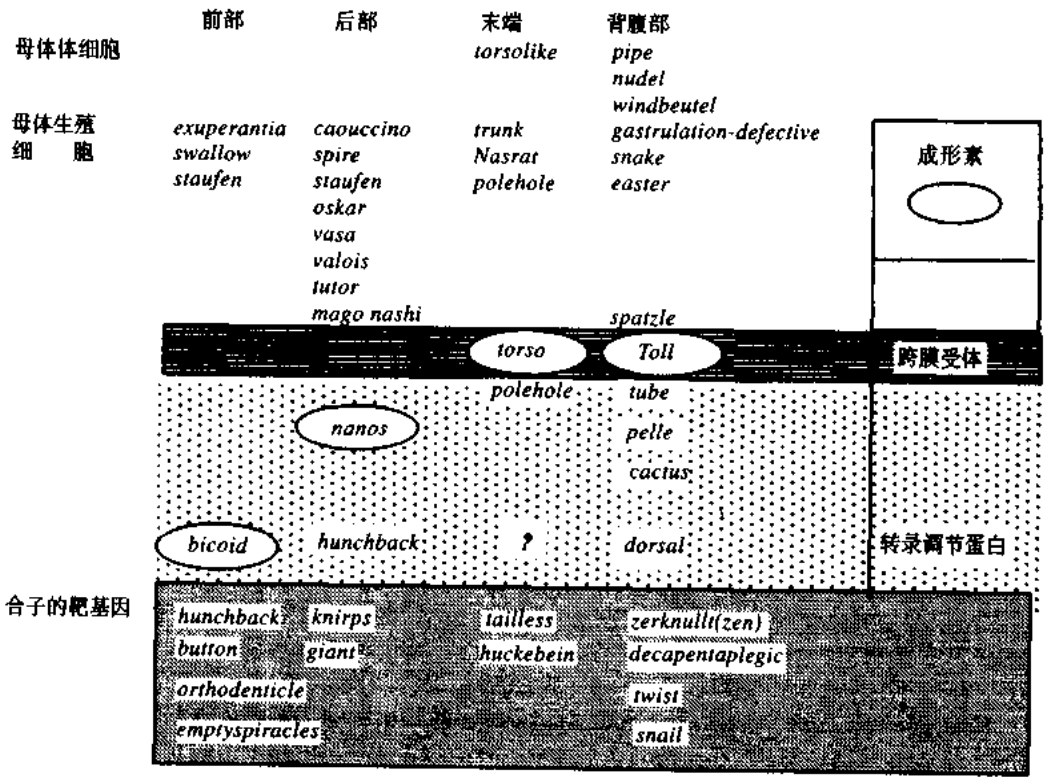


图 24 - 15 在卵中有四个母体系统发挥功能，其中每一个系统都是是在卵外合成，然后转运到卵中。进入卵的各种途径都具有一种定位产物，即形成素。这种形成素可能是一种受体或为基因表达的调节物，决定性的成分是转录因子。它们作用对于下一步发育作出反应的合子靶基因。(仿 B. Lewin: Genes, VI, 1997, Fig. 19 - 16)

这些途径的成分如图 24 - 15 所示。它们的操作有一个共同的原理，每个途径都是由所限区域而起始的，区域定位可以发生在卵中，也可在其外侧。这是由卵中的定位信

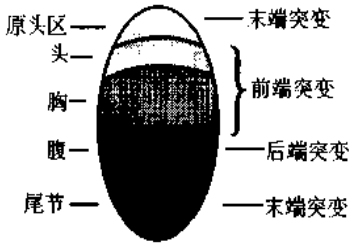


图 24 – 16　坐标基因的分布。

息所决定。这种信息是一种不对称分布的蛋白，称为**成形素**(morphogen)，这种蛋白在不同区域其浓度或活性不同，结果使其周围产生特殊的结构或命运。在各个这样的系统形成素或是一个转录调节因子，或导致所在区域转录因子激活。有 3 个系统是和前后轴有关；有一个系统和背复轴有关(图 24 – 16)。

一、前部系统

前部系统(anterior system)负责头和胸的发育。在此组中关键的基因是 *Bicoid*。此基因的突变导致胚胎缺失头和胸部，偶尔在前端有极性倒转的重复腹部体节。*Bicoid* 突变体的表型和经过某种手术处理所产生的表型相似。例如当果蝇的卵被钻一个小孔，让少量的胞质流出。若从前端失去细胞质的话，胚胎的前部发育成后部的结构；同样若采用注射的方法用后部的细胞质来取代前部的胞质，所产生的胚胎有两个镜像对称的腹部而没有头。此称做**异位表达**(ectopic expression)。

Bicoid 基因产物是一种序列特异 DNA 结合蛋白，是决定前部结构的转录因子。由于 *Bicoid* mRNA 位于早期卵裂胚的前部，因而前部的一些基因被激活。*Bicoid* mRNA 在滋养细胞中产生，然后通过胞质间桥转运到卵母细胞的前部区。其定位依赖于 3′端非翻译区。在合胞体卵裂时它形成了一种浓度梯度。胚的前部是浓度最高点。这种 *Bicoid* 基因的产物是决定胚胎命运的主要决定素。受 *Bicoid* 蛋白作用的基因含有多拷贝 9nt 组成的上游功能区 5 – TCTAATCCC – 3′。结合位点仅和保守序列相差 2 个碱基，此位点能以高亲和力和 *Bicoid* 的基因产物结合。若保守序列有两个碱基被替换，那么只能低效结合。高效和低效位点的重组决定了 Bicoid 蛋白结合的浓度，因此带有高效结合位点的基因能被低浓度的 Bicoid 蛋白所激活，而带有很多低效结合位点的基因就需要高浓度的 Bicoid 蛋白才能激活。这种不同的结合效率意味着基因表达的水平是由其结合位点的序列以及在前后轴上的位点所决定的。在母体中 *Bicoid* 基因拷贝数的改变也能改变这种梯度。这种梯度提供了效应基因表达的特异开关。*Bicoid* 对前部发育起到了一种指导作用。

Bicoid 蛋白激活的重要基因是 *Gap* 系列基因中的驼背基因(*hunchback*)。另外 6 个前部系统的基因编码 *Bicoid* 定位所必须的细胞组分。

二、后部系统

后部系统(posterior system)是负责腹部分节的一组基因。在腹节之后，其突变也能产生失去极性的细胞，一种后部突变称为 *nanos*，其产生的胚胎腹节有缺陷，但极性正常。突变体的表型与用手术去除后部胞质的结果相似。*nanos* mRNA 位于卵母细胞的后极，而此基因的产物是一个转录的阻遏物。在这些基因中，由于 nanos 的存在，*hunchback* 的 mRNA 不能翻译成蛋白。即 *hunckback* 受到 *Bicoid* 和 *nanos* 产物的共同调控。*Bicoid* 在前后的梯度中激活转录，而 Nanos 蛋白在后部区域阻遏转录。

三、末端系统

末端系统(terminal system)是负责卵不分节的最前部(吻部)和最后部(尾节)的发育。已发现了 7 个基因,关键的基因是 *torso*,它编码一个跨膜受体,在早期发育阶段均匀地分配到胚的各部分。此 Torso 受体仅在滋养细胞附近的卵细胞的极部被释放的信息激活。它是一种 Tyr 激酶,通过使靶蛋白中的特殊 Tyr 残基的磷酸化来启动细胞的分化。这种 Tyr 激酶其中有一个和脊椎动物原癌基因 *D-raf* 同源。

除决定胚前后轴的 3 套基因之外还有决定背腹(dorsal-ventral)的 4 套基因。这一途径是由来自卵腹部的滤泡细胞的一种信号而起始的。此信号是靠 *Toll* 编码的跨膜受体蛋白来传递。通过调控 *dorsal* 产物在细胞中的定位,使 *dosral* 产物的活性从腹部到背部形成一种梯度。Dorsal 蛋白是和鸟类原癌基因 *v-rel* 产物同源的转录因子。另外还有 16 个已知的基因影响背腹决定,这些基因的突变导致突变体失去腹部和晚期的模式因子。在很多情况下,这种突变的胚胎可通过注入野生型的细胞质而得到恢复。野生型的细胞质无论从什么地方取出,无论注入突变体的何处都能获得相同的效果。在背腹系统中 *snake*, *gastrulation-defective*(原肠形成一缺失)和 *easter* 三个基因编码丝氨酸蛋白酶(serine proteases)。随着前体的激活产生了 蛋白,而前体的激活又需要特异的卵裂的激活。也就是说在发育途径中一个酶的激活需要另一种酶的激活,出现了**级联效应**(cascade effect),从而放大了起始信号。

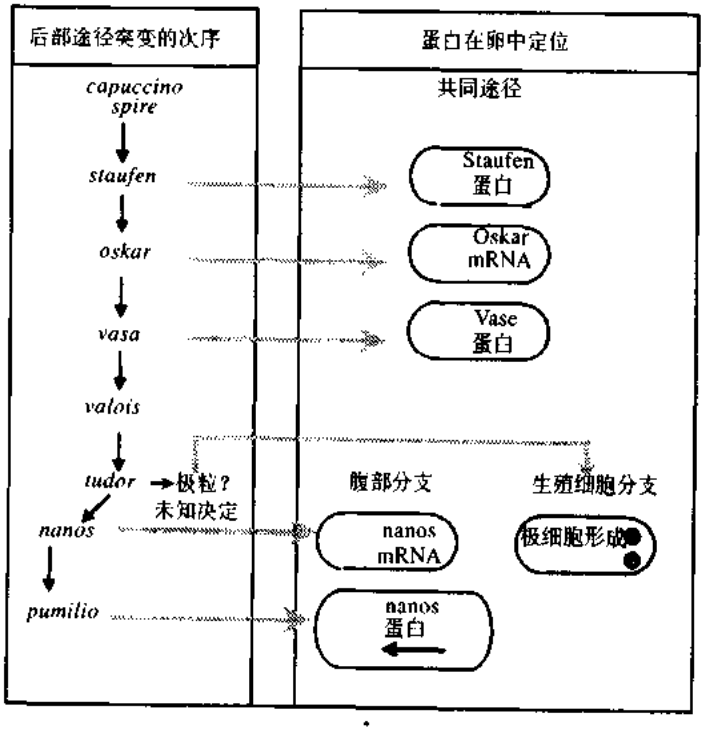


图 24-17 后部途径的潜在途径具有两个分支,它们分别负责腹部的发育和生殖细胞的发育。
(仿 B. Lewin: Genes VI, 1997, Fig. 38-7)

四、前后轴的发育

和模式形成有关的母体基因已被鉴别出 30 个左右，这 4 个系统中的所有成分都是母体基因，所以建立模式形成的起始是发生在受精前。身体的两条轴线（前后轴和背腹轴）是独立建立的。影响到极性的突变，使后部区域发育成前部结构，或者腹部区域发育成背部形态。按前后轴，前部和后部系统在胚的前端和后端提供了相反的信息源的梯度，从而控制了体节的发育。这两个系统中的任一个发生缺陷都会影响到体节的正常分化。末端系统和背腹轴系统独立地操纵着另一些系统。

通过**获救技术**（细胞质的抽提与注入）发现后部途径中有一种产物负责下一阶段的产物定位，图 24 - 17 表示在此遗传途径中，胚胎中的一些基因作用的相应次序。*spir* 和 *capu* 是 Staufen 蛋白在极区定位所必须的。Staufen 蛋白按一定的次序定位 *oskar* RNA。有可能 Staufen 蛋白和 *oskar* RNA 装配成复合体，此是 *vasa* 定位所必须的。Vasa 是一种 RNA 结合蛋白，其特异性和作用的靶是什么都还不清楚。

若 *oskar* 在胚中过量表达或错误定位都可导致在异位形成生殖细胞。它只需要 *vasa* 和 *tudor* 的产物，表明在此途径中排在 *oskar* 前的各种基因只起到定位 *oskar* RNA 的作用。也就是说与 *vasa* 和 *tudor* 相连接的 *oskar* 具有形成极细胞和腹部结构的能力。*oskar* 的功能之一就是将 *Vasa* 蛋白定位在胚的后端。*valois* 和 *tudor* 的功能还不清楚。

在后部极区发生两类模式决定事件，而在 *tudor* 后部途径产生分叉。极质含有两种形成素：后部决定因子（Nanos），控制腹部的发育；另一个是尚未鉴别的信号控制着极细胞的形成。极质区域将来发育为生殖系统。除了 *nanos* 和 *pumiho* 以外其余的 6 个基因都是腹部发育和极细胞的形成所需要的。*nanos* 和 *pumiho* 参与腹部的分化。

前部系统和后部系统形态发生的基本特点是相似的。母体的 mRNA 被定位在后部极区。这就是 *nanos* 的产物，它提供了一种成形素。这两个系统有两个重要的区别是后部系统的定位要比前部系统复杂。这是由于后部决定因子是在滋养细胞中形成的，它必须经过转运才能进入卵母细胞到达远处的极区，而 Nanos 蛋白作用是抑制转录因子（*hunchback*）的翻译，其功能是抑制那些影响后部发育的基因。

用原位分子杂交表明 *nanos* 的 mRNA 位于早期胚的后端但它的功能是控制腹部的发育。这是由于 Nanos 蛋白从后端延着腹部区域向外扩散，形成了一定的浓度梯度。后部的外侧对 *nanos* mRNA 的翻译具有抑制作用。*nanos* mRNA 的定位和抑制都取决于其 3′端区域的翻译序列。

前面已提及 *bicoid* 和 *nanos* 影响 *hurchback* 基因的表达。*hurchback* 编码一种转录抑制因子。它的存在对于前部结构的形成是必需的（胸部区域），而它的缺乏对于后部结构的发育也是必需的。它具有一种复杂的表达模式——它在卵子发生时被转录，在卵中产生均匀分布的 mRNA，受精后 *hunchback* 模式以两种方式发生改变；Bicod 梯度激活 *hunchback* mRNA 在前部区域的合成；而 Nanos 在后部区域抑制 *hunchback* mRNA 的翻译，结果导致了这种 mRNA 的降解。因此 *hunchback* 的水平在卵前半部分增强，而后半部分被消除，这种分布的重要性就在于 Hunchback 的调节作用，它可抑制形成腹部结构所需要的 *knirps* 和 *gicnt*。这样 Hunchback 的基本作用是通过抑制 *kmirps* 和 *giant* 的表达，从而阻止形成腹部结构。

五、背腹轴的发育

背腹发育展示了卵母细胞和滤泡细胞之间十分复杂的相互作用。在卵母细胞中模式形成开始时一些基因的表达对于滤泡细胞的发育是必要的。而在滤泡细胞中一些基因的表达又可将信号传递到卵母细胞中，结果导致了腹部结构的发育。前部途经是负责胚的背部发育。每个系统是通过激活已定位的配体受体相互作用，触发信息传导途经来实现它们的功能。

这个过程一开始 *gurken* mRNA 进行定位，此分为两个阶段，首先前后模式起作用，然后实施背腹模式，此事件的关键是卵腔空间的对称性，*gurken* 对于两个系统来说都是需要的，其特点是具有独立的活性。

首先 *gurken* mRNA 定位在卵母细胞的后部，这样它作为一种信号促进邻接的滤泡细胞变成胚的后部。这种滤泡细胞的信号又反馈到卵母细胞，导致形成了微管的极化网，这对于 *bicoid* 和 *oskarin* 的转录定位和趋向两极是必要的。

背腹极性在 *gurken* mRNA 定位在卵母细胞的背侧之后得以建立。图 24－18 表明了这个途经及其结果。*cornichon*（*cni*）和 *brainiac*（*brin*）的产物是 *gurken*（*grk*）mRNA 正确定位或该蛋白的激活所必须的。早期起作用的一组基因中，K10 和 *squid*（*sqd*）的产物对于 *cri* 和 *brin* RNA 的定位是必须的。而 *cappuccino*（*capu*）和 *spire*（*spir*）突变体产生了排列缺陷，表明它们的产物在卵母细胞的细胞骨架组成中发挥作用。这两个基因对于和前后模式有关系的 *gurken* mRNA 早期定位也是必要的。

gurken 编码的蛋白和生长因子 TGF_2 相似。在此途经中的下一个基因座位是 *torpedo*（*top*），它编码果蝇的 EGF 受体，在滤泡细胞中表达。配基－Gurken 以跨膜的形式延伸到卵母细胞的胞外区，与滤泡细胞质膜上的受体 Torpedo 相互作用，结果这个途经从卵母细胞移到滤泡细胞。

作为典型的受体 Tyr 激酶 Torpedo 的激活体现了一个有趣而普遍的原理。Torpedo 的激活导致 Ras 信号途经的激活，此途经通过 Raf 和 D-mek（相当于 MAPAA），激活经典的 MAP 激酶途经。这个途径的终点尚不清楚，但

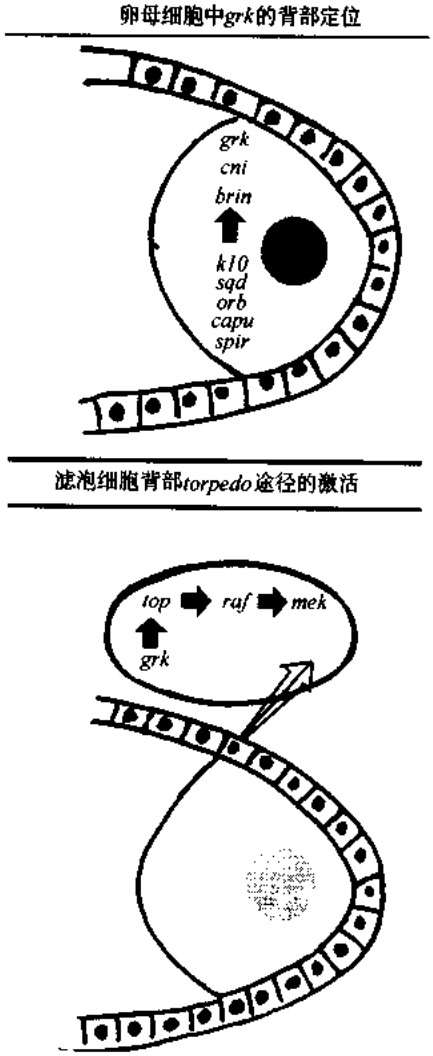


图 24－18 背部和腹部的同一性是当 *grk* 定位于卵母细胞的背侧时首先被识别，*grk* 的合成激活了 *torpedo* 编码的受体，在滤泡细胞中触发了 MAPK 途径。
（仿 B. Lewin: Genes VI, 1997, Fig. 38－9）

它的作用阻止了下面要讨论的腹部决定途经的胚胎背部的激活。

这个途径的利用表明相似的途径可以用于不同的情况来产生高度特异的效应。在卵母－滤泡细胞中激活此途径的触发点是 Gurken 的特殊定位。结果是改变滤泡细胞的性能，阻止它们获得腹部命运。这个途径的主要成份是和用于脊椎系统中信号

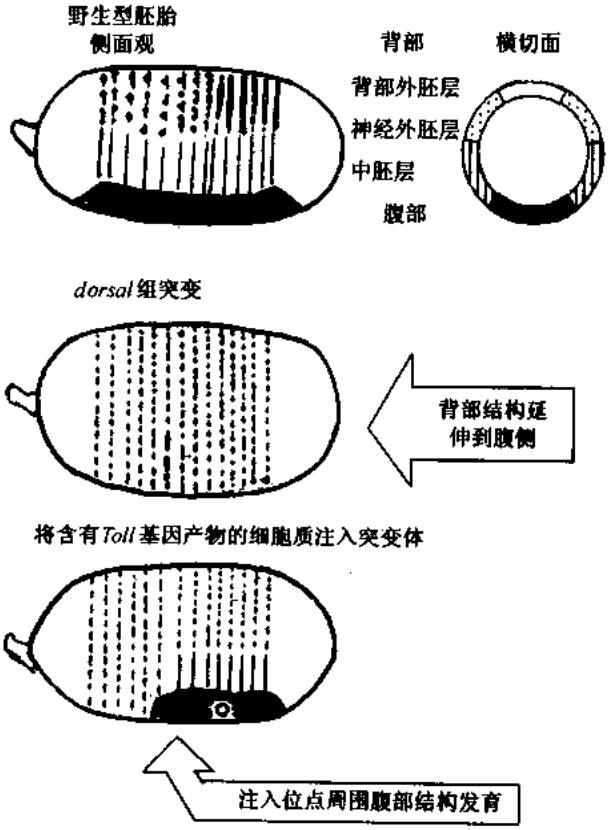


图 24 － 19　野生型早晚胚胎具有不同的背部和腹部结构。*dorsal* 组基因突变抑制了腹部结构的发育，腹部产生了背部结构，通过注入含有 *Toll* 基因产物的细胞质腹部结构能得到恢复。（仿 B. Lewin: Genes Ⅵ，1997，Fig. 38 － 9）

传递的成分相同，实际上也用于果蝇视网膜特殊发育中，那儿的另一些受体－反受体之间（receptor-counter receptor）相互作用激活 Ras 途径，特殊的是它对不同的细胞起完全不同的作用。这样基本相似的途径能用于内部信号，并产生由细胞表型事先决定的反应。

腹部结构的发育需要一组 11 个母体基因，它们的产物在受精和细胞囊胚形成之间这段时间建立背腹轴。图 24 － 19 表明通过表皮表型的侧面观来了解背腹模式。从横切面可以观察到不同组织的形成。*dorsal* 系统控制腹部发育，包括中胚层和神经外胚层的发育。这个系统的命名是由于其突变的效应是“背部化”而不是“腹部化”。*dorsal* 组的很多若基因发生突变将会使腹部不能产生其特有的结构，而在腹侧形成背部结构。但若将野生型胚胎的细胞质注入到突变胚胎可使腹部结构正常发育。

腹部决定途径也是从滤泡细胞和卵母细胞的末端开始。在图 24 － 20 表明了此途径。开始阶段尚未很好地限定，在腹部的滤泡细胞中需要有三个座位的表达。这些座位的功能在受精前就存在，但卵在受精前不能接受这些信号。虽然我们已经知

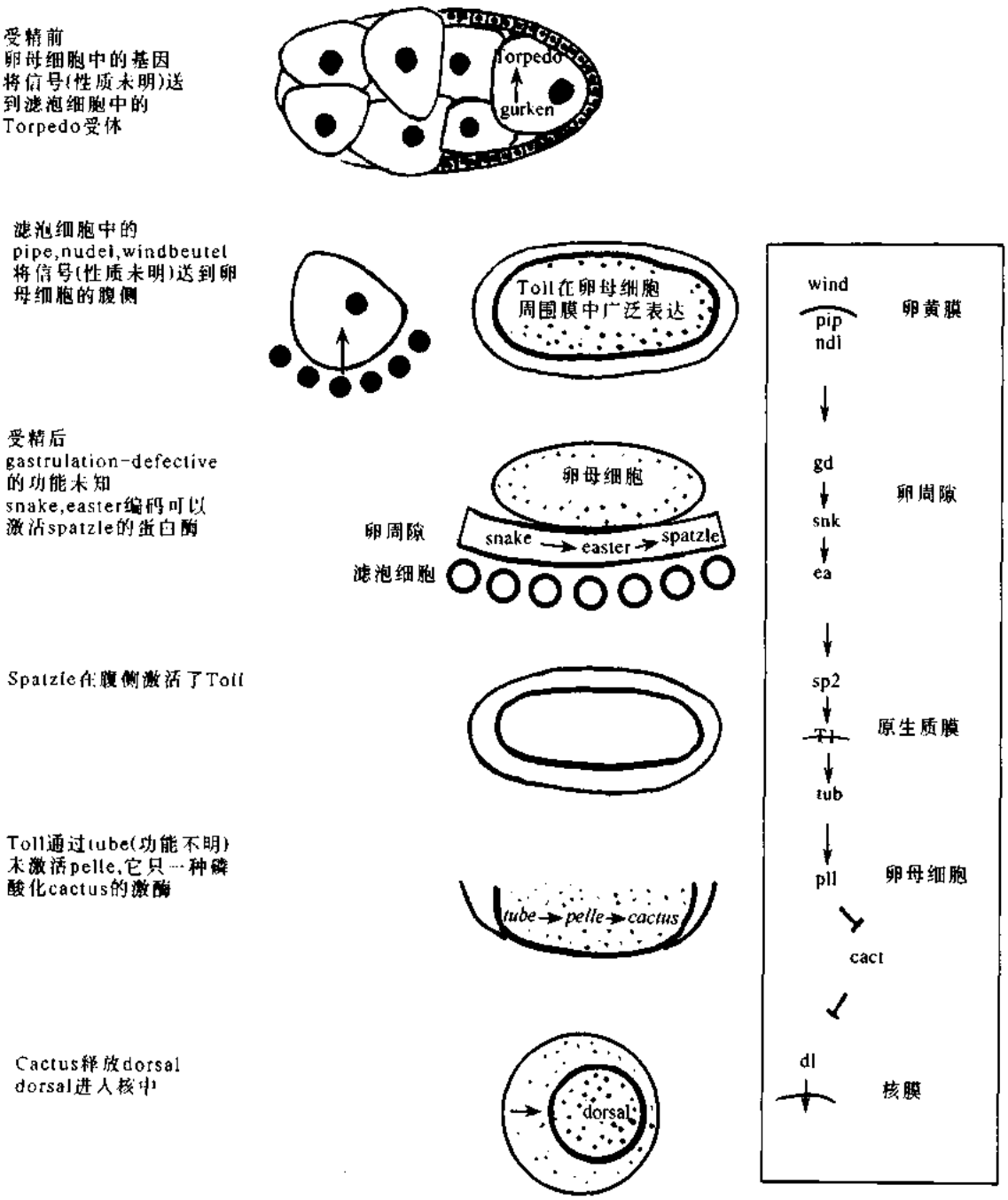


图 24 - 20 dorsal - 腹部途径涉及滤泡细胞和卵母细胞之间的相互作用。当 spatzle 结合于 Toll 并激活形态形成素时此途径移入卵母细胞中。此途径是通过转录因子 dorsal 转运入核中而完成。
(仿 B. Lewin: Genes VI, 1997, . Fig. 38 - 11)

道在卵母细胞中随后发生的大部分事件，但还不知道这种信号的性质。

此信号在**卵周隙**（ perivitelline space ）（卵母细胞的外层）中导致发生一系列的蛋白水解剪切。可能由 *snake* 的产物剪切 *easter* 的产物，依次被激活，剪切 *spatzle* 产物。 *spatzle* 被剪切激活，因此它能为由 *Toll* 基因编码的受体提供配基。 *Toll* 是卵母细胞中主要途经的第一个成分。

获救实验表明 *Toll* 是一个将信号传入卵母细胞中的重要的基因。 *Toll* - 突变体失去所有背腹梯度，但注入 Toll 就会引起背腹结构的形成。背部的另一个基因编码

的产物或是调节物，或是为 Toll 的作用所需，但它们并不能建立初步的极性。在 Toll 蛋白的分布中存在一些矛盾。当进行细胞质抽提和注入时，供体胚胎的各部分都发现 *Toll* 基因产生的活性。而在正常发育中只有在适当定位的情况下它才引起腹部结构的形成。*Toll* 基因产物在内部的一般分布必须以某种方式通过定位转变成浓缩的活性产物。

Toll 是一种与脊椎动物白细胞介素 - 1 受体同源的跨膜蛋白。它的作用模式尚不清楚，但它与配基结合后就能有效地激活腹部决定途径。这个反应发生在卵周隙的腹侧面。Spatzle 作为其配基，它的分布不能离剪切位点太远，或者它可以迅速和 Toll 结合，结果 Toll 仅在胚胎的腹侧被激活。经突变失去了功能的 Toll 使胚胎的腹部背部化，这是由于受体不能被激活。也有一种显性突变（$Toll^D$），它使得背部出现腹部的特征；这些是**功能获得**（gain-of-function）突变，它使得胚胎腹部化，这是由于受体被持续活化。不论用何种方法 *Toll* 通过 *tube* 和 *pelle* 起作用。*tube* 的功能还不知道，*pelle* 编码一种激酶，这种激酶的靶位点可能是 *cactus* 的产物。Cactus 又是 *dorsal* 编码的转录作用因子的调节物（它有可能也像 Pelle 激酶一样，直接作用在 Dorsal 蛋白上）。

Dorsal 和 Cactus 形成了一对相互作用的蛋白，它们与转录因子 NF-κB 和它的调节蛋白 IκB 相关联。NF-κB 含有 2 个亚基（序列相关）在胞质中通过 IκB 它们相互结合。当 IκB 被磷酸化时，它释放 NF-κB ，进而进入核中，在那儿 NF-κB 作用具有 κB 序列基序启动子的转录因子。看来 Cactus 调节 Dorsal 的方法和 IκB 调节 NF-κB 相似，即 Cactus-Dorsal 复合体在胞质中是无活性的，但当 Cactus 被磷酸化后它释放了 Dorsal 蛋白，Dorsal 再进入核中。从受体到效应物的途径是保守的，白细胞介素 -1 受体的激活具有和 NF-κB 激活相同的主要作用。

Toll 的激活使得从胚胎的腹部到背部建立了 Dorsal 蛋白的梯度。在腹侧 Dorsal 蛋白释放到核中，但在胚胎的背部却留在胞质中。陡峭的梯度是在合胞体阶段建立的。在变为细胞囊胚时成形。Dorsal 蛋白的特点是仅在腹部分布。在胚胎中 Dorsal 蛋白的总量不改变，而是通过核和胞质之间的蛋白再分配建立这种梯度。

Dorsal 能激活和抑制某些基因的表达。它可激活腹部结构发育所需的 *twist* 和 *snail* 基因，而可以抑制背部发育所需的 *dpp* 和 *zen* 基因。结果这些基因在胚胎的背部只有 40 % 可以表达。

dpp 的产物是 TGFβ（转化生长因子 β）生长因子家族的一个成员。由于此基因在腹侧被抑制，所以 dpp 蛋白仅从胚胎的背侧细胞分泌出来。这样 *dpp* 实际上是指导背部结构合成的形态成形素。这个家族成员的受体是典型的具有 Ser/Thr 激酶活性的异二聚体跨膜蛋白。此异二聚体是由Ⅰ型和Ⅱ型成分构成。在 *dpp* 途径中有两种Ⅰ型成员（由 *thick veins* 和 *saxophone* 编码）和单个的Ⅱ型成员（由 *punt* 编码）。Tkv/Punt 二聚体是基本的受体。在下游途径中配基/受体相互作用的特点还没有搞清楚，尽管已知 Zen(由 zerknult 编码)转录因子是其中的成分之一。

很多座位通过大量翻译后机制干扰 dpp 的产生。这种影响的结果是在胚胎的背侧增加 dpp 的活性，而在腹侧抑制其活性。dpp 的浓度直接影响了细胞的表型，大部分背部的表型是取决于 dpp 的浓度梯度。

胚胎背腹部发育的决定性因素之一是定位系统之间的关系(图 24 - 21)。一个系统对下一步的抑制能力是负责限制胚胎适当部分的定位活性。在 gurken 和 torpedo 之间相互作用的起始导致了胚胎背侧 Spatzle 活性的抑制。这样就限制了胚胎腹侧 Dor-

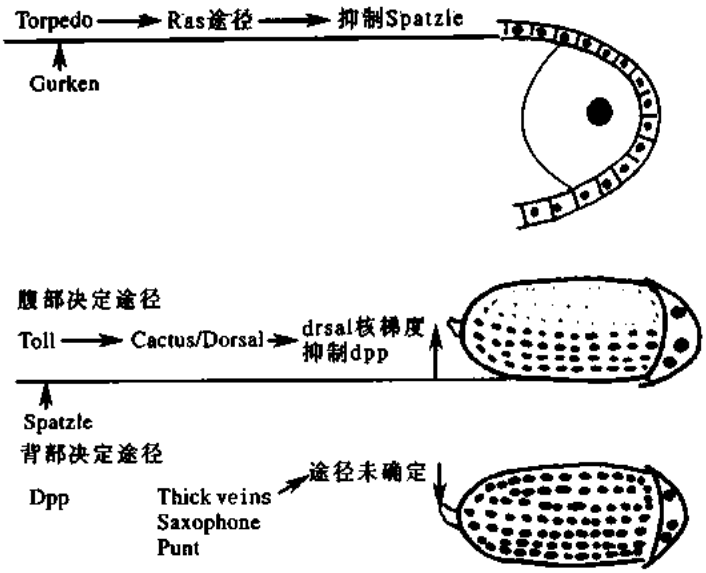


图 24－21 背腹模式需要 3 个定位系统连续的活性。
（仿 B.Lewin:Genes VI,1997, Fig.38－13）

sal 蛋白的活性。Dorsal 蛋白的定位依次抑制 *dpp* 的表达，这样它形成了从背侧散开的梯度。以这种方式腹部结构在 Dorsal 蛋白的梯度中形成，而背部结构在 Dpp 蛋白梯度中形成。

末端系统起始的方式是和背腹系统相似的。由 *torso* 基因编码一个跨膜受体，

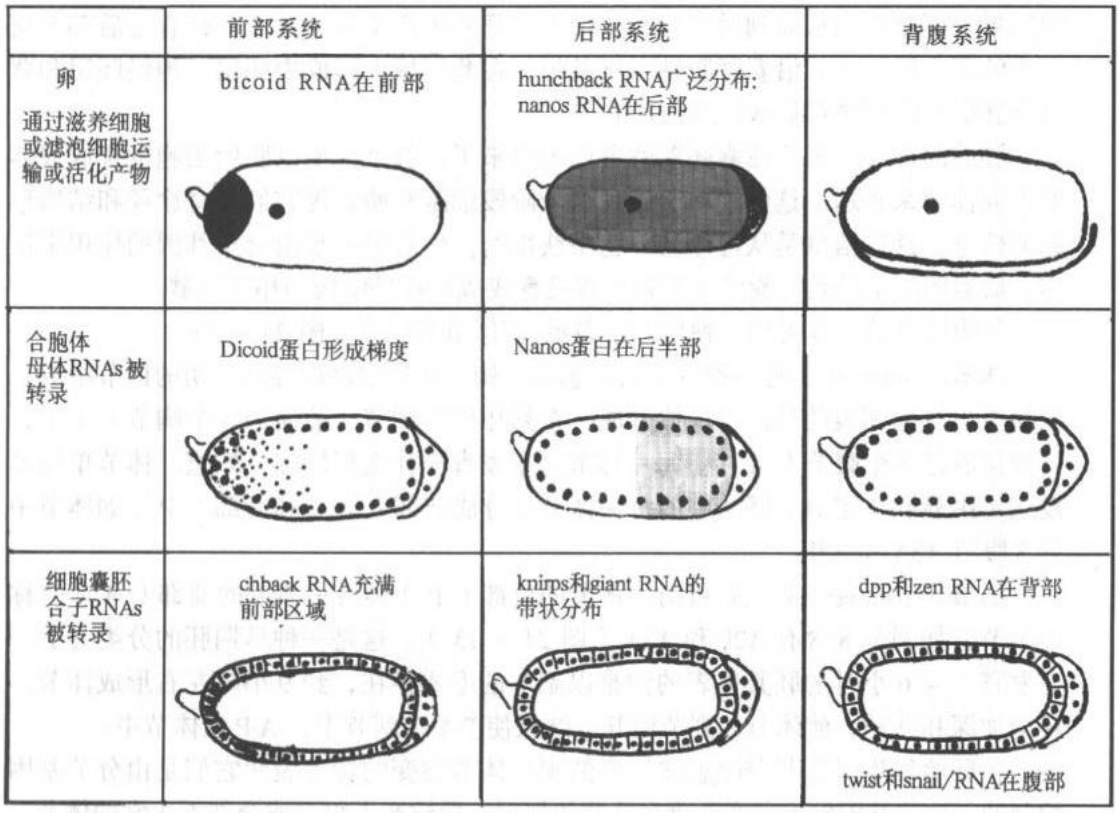


图 24－22 在每个轴一决定系统中，卵内的产物定位导致另一些母体 RNA 或蛋白在合胞体中的广泛定位以及合子 RNAs 在细胞囊胚中广泛地转录。（仿 B.Lewin:Genes VI,1997, Fig.38－14）

它是母体 RNA 在受精后翻译而成。此受体定位遍及整个胚胎，通过胞外配基的局部产生而被激活。Torso 蛋白具有起始级联反应的激酶活性。这种级联反应导致了 *tailless* 和 *huckebein* RNAs 的表达。这两个基因负责编码一种调节转录的因子。

对于各个系统而言在每个阶段的调节模式如图 24 - 22 所示。不对称的起始有两种机制。对于前后轴来说，RNA 在卵的一端定位（前部系统的如 biclid，后部系统的如 nanos），此定位是依赖于 RNA 3′端的序列与母体蛋白的相互作用。在背腹系统和末端系统中，由于配基的作用受到了限制，使受体蛋白以定位的方式被特异的活化，所有的这些相互作用都取决于由母体基因表达产生的 RNAs 和/或蛋白。

局部的事件导致产生了形态成形素，它或是形成数量的梯度（如 *Bicoid*），或是通过核质分布（如 dorsal），或是定位于一定的范围区域内。在这个区域范围内形态成形素被激活，对于各系统来说这种区域范围越过卵的一半。形态成形素是母体的 RNAs 转录而成，因此直到这个阶段发育仍然依赖于母体基因。

第五节　分节基因的调节作用

在胚胎的定向决定和空间定位中第一步就是建立前后轴梯度和背腹轴梯度。在母体基因的指导下，梯度贯穿整个胞质，而且会影响胞质中核的行为。第二步决定身体不同区域的发育。这一步是需要合子基因组的表达，根据它们的活性现在把这些基因称为**合子基因**（zygotic gene）。和此阶段有关的基因是通过分节突变被鉴定出来的。

分节基因的产物沿着前后轴形成不同的分布区。这是前部和后部系统共同作用的结果。不难看出从前到后形成几条带（前部系统产生两个区域，接着是后部系统产生的两个区域）。沿着背腹轴有三条带，在此区域中形成中胚层、神经胚层和背部外胚层（排列序列是从腹部到背部）。

胚胎的“命运图”在囊胚期就可以画出来了，各个区域以胚胎细胞发育成成体的不同体节来表示。这可能由于通过这个阶段细胞开始获得它们发育途径和结构形态的信息。这些信息是从母体调节物中获得的，然后进一步由合子基因的作用来证实。命运图的本质是在囊胚上鉴别出将发育成成体特异结构的相应区域。

黑腹果蝇的发育可用二种单位来表示，即体节和副节（图 24 - 13）。

体节（segment）是一种可见的形态学结构。成体果蝇由界线分明的体节组成，而幼虫也有由体表的环沟将身体分成一系列相应的体节。前面为 3 个胸节（T），后面接着是 8 个腹节（A）。每一体节将来发育成什么部位都已知道。体节单位的模式是由囊胚决定的。胚胎中的每一体节又分成前部（A）和后部（P）如体节中第 3 腹节 A3A 和 A3P。

副节（parasegment）是将前一体节的后部（P）和后一体节的前部（A）合称为副节，如副节 8 含有 A2P 和 A3A（图 24 - 13）。这是一种早期胚的分类方法。在发育 5 ~ 6 小时的胚其体表的沟是以副节的形式存在，到 9 小时左右形成体节。环沟加深和移动，使体节和副节错开。结果使 P-A 在副节中，A-P 在体节中。

在胚胎发生时这些分隔怎样形成的呢？体节突变的特点表明它们是由分节基因控制的。这些基因发生突变会改变分节的规则，导致产生很多或全部不正常的体节。分节畸型的严重后果是使突变体胚胎致死，一般在变成为成体之前的各个发育阶段中死亡。可能约有 30 个座位涉及到体节形成。图 24 - 23 表明根据它们作用单位的

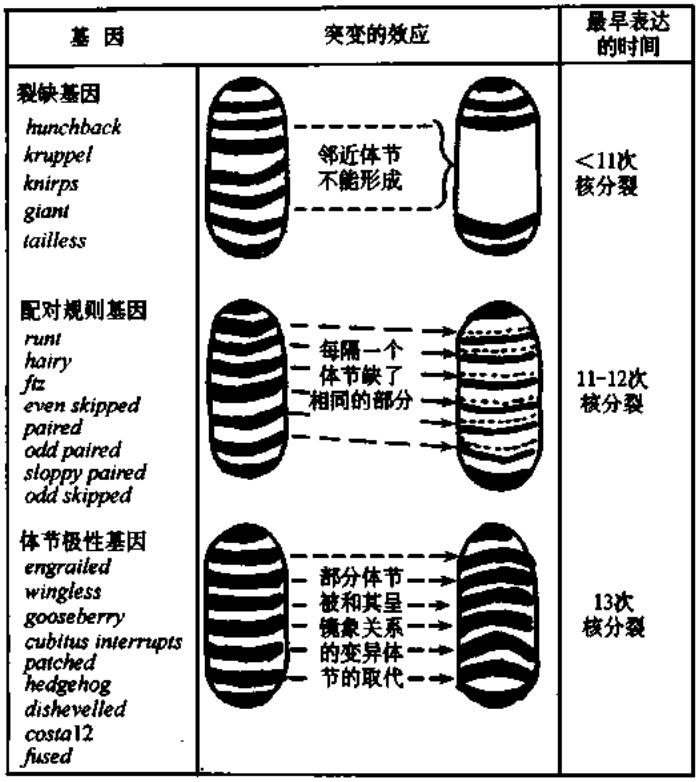


图 24－23 分节基因影响到很多体节。相关基因可分为 3 组。
（仿 B. Lewin: Genes VI, 1997, Fig. 38－16）

大小进行的分类。

裂缺基因（*gap*）在胚中的一些体节中表达，建立空间结构的相邻平面，*gap* 突变有一组体节缺失。有 4 个 *gap* 基因和身体主要体节形成有关，其他的和头部及尾部结构有关。

配对规则基因（*pair-rule*）使胚胎分成为明显的体节，这组基因的突变体每隔一个体节就缺失了一部分。所涉及的体节可能是偶数的，也可能是奇数的。这一组有 8 个配对规则基因。这些基因延着胚胎的前后轴以狭窄的斑马条纹的形式表达。

体节极性基因（*segment polarity*）决定胚胎每个体节的发育形式。其突变体常常丢失了每个体节的 P 成分而由被其呈镜像重复的 A 成分所取代。有的突变体丢失了A成分或是中部体节。这组基因有 16 个。

这 3 组基因在发育中相继表达，它们在卵中形成的限制区域逐步增加（图 24－24）。母体基因从前、后端建立了梯度。

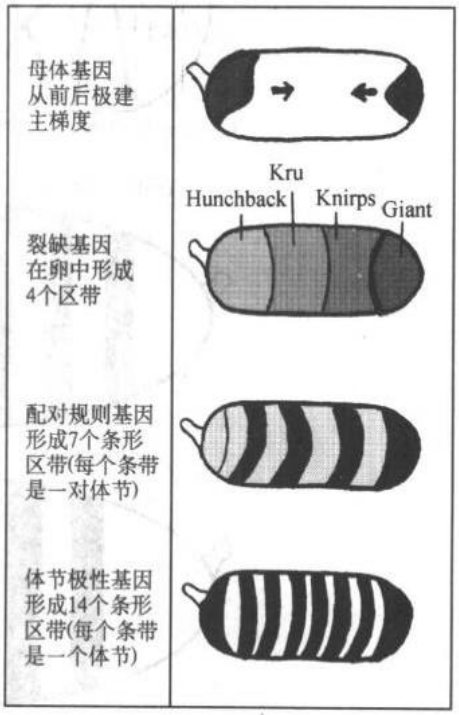


图 24－24 母体基因和分节基因逐渐地作用在胚胎的相同区域上。（仿 B. Lewin: Genes VI, 1997, Fig. 38－17）

这种梯度激活或抑制了裂缺基因。裂缺基因随着受精（在第 11 次核分裂时）最早被转录；它们将胚胎分成 4 个区带。裂缺基因调节稍晚一点转录的配对规则基因。它们的靶区域被限制为成对的体节。配对基因依次调节在第 13 次核分裂时表达的体节极性基因，配对规则基因的靶位点是单个的体节。

很多的母体基因，裂缺基因和配对规则基因都是转录调节物。它们的作用可以是激活，也可以是抑制；在某些情况下特定的蛋白可以活化某种靶基因，而抑制另一些靶基因，取决于其浓度和竞争。这类基因不但相互调节，也调节下一组基因。当分节极性基因达到一定水平时，我们才能找到其产物，它们以其他的方式起作用；例如从一个细胞中分泌的蛋白会影响到其邻近的细胞。

在发育的各个阶段都显示出这样的规律：少量的母体基因、裂缺基因和配对规则基因的调节，在胚胎的特殊区域中经重组装配控制基因的特殊表达模式。

裂缺基因以两种方式受到调控：(1)它们可能直接对 Bicoid 形态形成素作出反应；(2)以及它们相互调节。图 24 - 24 中所显示 4 条区带是按下例方式产生的。最前面的带是由 Hunchback 蛋白组成的，此蛋白由 Bicoid 激活转录的 *hunchback* RNA 合成的。第二条带是由 Kruppel protein 构成的。而且 *Kruppel* 基因的转录是 Hunchback 蛋白激活的。后面的两条带是由 Knirps 和 Giant 蛋白构成。Hunchback 蛋白可抑制这两个基因的表达。这两个基因在后部表达是由于 Nanos 抑制了此处的 *hunchback* 的表达。

图 24 - 25 表示从 4 条带变成 7 条带的细节。各种裂缺蛋白相互作用的细节是通过在缺少一种特殊裂缺蛋白的突变体中检测另一些裂缺蛋白的分布模式而确定的。

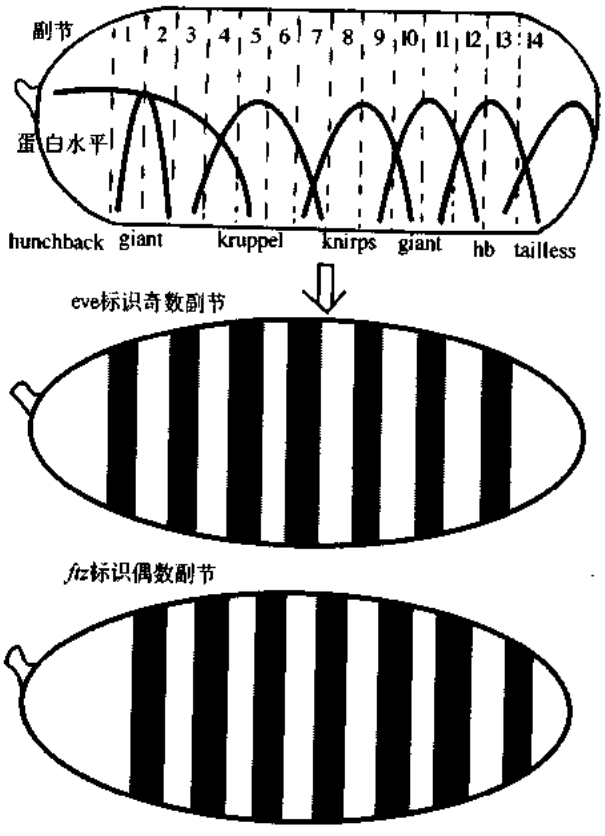


图 24 - 25　裂缺基因的表达限定了胚胎的邻接区域裂缺基因控制配对规则基因，其中每一个在 7 个条带区域中表达。（仿 B. Lewin: Genes VI, 1997, Fig. 38 - 18）

Hunchback 起到一种特别重要的作用。它在前部区带中表达（图 24 - 24），在胚胎的中部梯度下降（图 24 - 23），高水平的 Hunckback 抑制 Kruppel；这就决定了 *krup*pel 表达的前部分界线，其起始和上升处在第 3 副节，正好是 Hunchback 水平下降的地方。但 Hunckback 的一定水平对 *kruppel* 的表达是需要的，因此当 Hunckback 的水平在第 5 副节进一步减少时，*kruppel* 被关闭。Giant 以同样的方式对 Huckback 水平的改变作出反应。

通过这些产物的相互作用使调控更为精炼。一般的原理是在特殊区域中一个蛋白表达可能需要一种蛋白的相互作用，在分界线上抑制其表达可能需要另一种蛋白相互作用。这种效应是通过成对位置相互作用的实验而解决的。比如 *giant* 的过度表达会导致 *kruppel* 带变得非常窄，表明 *giant* 的分布抑制 *kruppel* 的分界线。*kinirps* 和 *giant* 的后部边缘是由末端系统的作用而确定的。这些相互作用意味着延着卵从前到后任何一个特殊位点都可能受到裂缺蛋白变化水平的限制。

所有的裂缺蛋白都受到转录调控以及附加的相互调控，它们又调节在下一阶段起作用的配对规则基因的表达。延着胚胎在 7 个窄条区中都可发现各种配对规则蛋白，图 24、25 表示这些窄条区的大约位置，裂缺基因表达的结果将会产生这些窄条。（当然这时副节还没有发育，仅表示这些蛋白分布的相对位置。配对规则基因的 7 条窄带或者作为所有奇数副节的标识（像 *eve*），或作为偶数副节的标识（像 *fte*）。*Hairy* 和 *eve* 这两个配对规则基因被称为**初始配对规则基因**，这是由于它们首先表达，而且其表达的方式影响了其他配对—规则基因的表达。

配对 - 规则基因的突变导致了一半体节缺失。图 24 - 26 表明野生型和 *ftz*(fushi-tarazu)突变型幼虫的比较；突变体仅有半数的体节，其他体节都失去了。

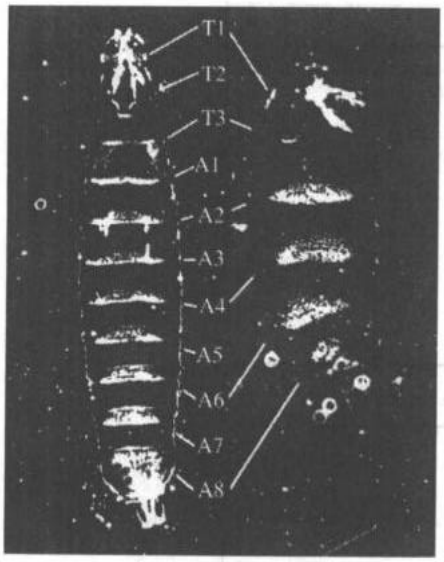


图 24 - 26 *ftz* 突变体的体节数只有野生型的一半。
（引自 B.Lewin:Genes VI,1997,Fig.38 - 19）

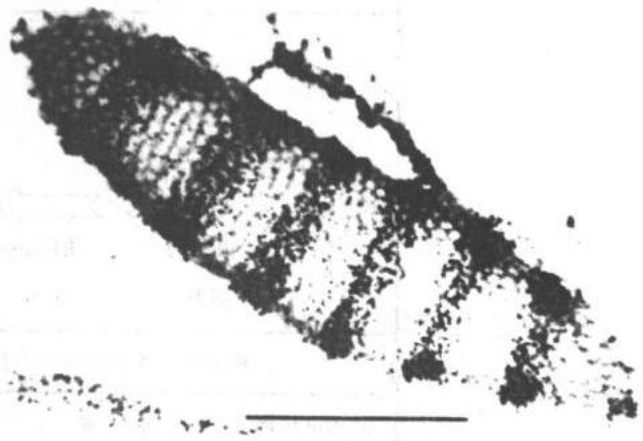


图 24 - 27 *ftz* 的转录物的定位条纹排在偶数副节上。图中野生型所表达的区域正是 *ftz* 突变体缺乏的区域。
（转引自 B.Lewin:Genes VI,1997,Fig.38 - 20）

从早期的囊胚到原肠胚都存在 *ftz* mRNA。图 24 - 27 表示转录本的定位。这是通过野生型囊胚的原位分子杂交获得的。此基因在 7 个窄条中表达，每条 3 ~ 4 细胞宽，横穿整个的胚。正如图 24 - 25 所示，这些窄条和偶数副节相对应（如 4 =

T1P/T2A， 6 = T3P/A1A， 8 = *A2P/A3A* 等）。

此模式表明 *ftz* 基因的功能：它在囊胚中表达对于形成由偶数体节发育而成的结构来说是必要的。*ftz* 的突变体失去了这些副节，这是由于在此期缺乏形成结构所需的该基因的产物。另一方面 *ftz* 的表达对于 *ftz* 所在细胞的存活也是必要的。

ftz 的表达是一个基因控制结构发育的例子，它表达所在的窄条区与其突变时胚胎失去的区域相一致。分隔是由分节基因的表达所决定的。基因表达所在的窄条区域宽度和它所影响的体节单位的大小相一致。不同的配对规则基因的表达模式其机制也不相同。人们已掌握了很多有关 *ftz* 和 *eve* 的信息。

在早期胚中 *ftz* 的表达没有什么变化的。若在窄条区域出现前停止了蛋白合成，那么胚胎将保持着开始的样子。也就是说窄条区域的发育依赖于在胚胎的带与前、后端之间区域 *ftz* RNA 的降解。一旦窄带区域已出现，则在带间和胚胎的两端 *ftz* 便停止转录。这种特异性转录依赖于 *ftz* 启动子的上游区域，同时也依赖于其他一系列分节基因。*ftz* 的转录通过作用于所有窄带区的因子对其他的配对－规则基因（和裂

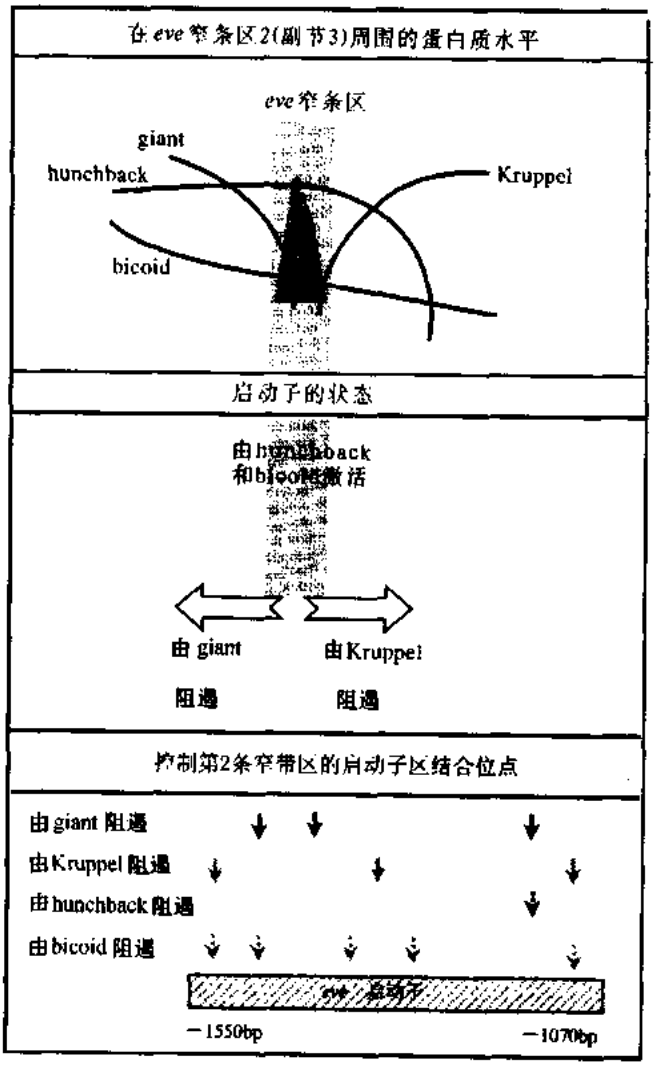


图 24 － 28　在副节 3 中的 eve 窄条区由 hunchback 和 bicoid 激活，由前部区带的 giant 和位于后部区带的 Kruppel 阻遏。在这个启动子的 480bp 的区域内有很多结构位点。控制着基因的表达。（仿 B. Lewin: Genes VI, 1997, Fig. 38 － 21）

缺基因）作出反应。

eve 的表达模式与 *ftz* 互补，但基础不同：它是在每个窄条区分别受到调控的。用 *eve* 启动子进行重建的实验表明在每个窄条区的定位信息是由启动子的不同部分编码的；其启动子可分为与在特殊副节裂缺基因产物局部水平相关的区域。例如在第 3 副节中负责 *eve* 表达的启动子区域具有裂缺蛋白 Bicoid，Hunckback，giant 和 Krup pel 的结合位点。图 24－28 表示其启动子 480bp 的这一区域及调控的途径：*eve* 转录是由 hunckback 和 Bicoid 激活的，由于它受到前部的 Giant 和后部的 kruppel 的阻遏（图 24－25）而确定了两个分界线。启动子的另一部分对胚另一部分的蛋白质水平作出反应。初始配对－规则基因产物的不同窄条区受到不同途径的调控，每一窄条区对于裂缺基因的产物以及其他调节物特殊组合的激活是敏感的。

图 24－28 表明这样的一个原理：蛋白综合调控局部区域基因的表达。一般的原理是蛋白的广泛分布（例如：*bicoid* 或 *hunckback*）对于激活是必要的，节段的分界线是由抑制形成的（在此特例中是通过 Giant 和 kruppel 进行抑制的）。在此必须强调基因调节的等级化并不排除连续调节（母体基因 裂缺基因 配对规则基因）之间有限的相互作用；例如在副节 3 中调节 *eve* 转录的 Bicoid 蛋白表明母体基因对配对－规则基因可能有直接的作用。

图 24－29 *ftz*（褐色）和 *eve*（暗灰）经同时染色表明它们在囊胚期首先表达成交替排列的宽带（上图）在下一个小时的发育中带变窄（下图）。（引自 B. Lewin：Genes VI，1997，Fig. 38－22）

eve 和 *ftz* 的窄条开始是比较模糊的，在发育的过程中突然变得很清楚。图 24－28 表明 *ftz* 和 *eve* 表达经同时染色的胚胎，开始出现一些交替模糊窄带，这些窄带离后端很近，而在发育时前端的带突然加深，这可能取决于自我调节循环，在此调节中基因的表达是由它们自身的产物来调节的。

配对－规则基因控制分节极性基因在第 14 窄条区的表达。每个窄条区等同于一个体节。体节极性基因表达的分隔方式精确。精确性的最基本的论证是由 *engrailed* 基因提供的。*engrailed* 编码的功能是所有体节所需的，而且和 A、P 分隔之间的差别有关。engrailed 蛋白的抗体和它所在细胞的核发生免疫反应，Engrailed 表达的区域形成了一种窄条的式样。当 Engrailed 蛋白的窄条区首先变得明显，它们只有一个细胞带。图 24－29 表示每个体节有一条约一个细胞宽的窄带，然后加宽到几个细胞。

实际上窄带区的形状是在 30 分钟后建立的，并从胚胎的前面向后面移动。开始的第一条窄带是明显的；然后每一体节都有一个窄带。最终每个体节与 P 分隔相应的地方都有一条 3～4 细胞宽的窄带。

engrailed 的表达是特别重要的，因为它决定了实际区室的分界线，以后将发育为成体的结构。开始的一个细胞宽的 Engrailed 蛋白窄带在 *ftz* 和 *eve* 窄带的前部分界线上形成并显示出前部每一个 P 区室的界线。为什么 *engrailed* 开始时只在前缘 *ftz* 和 *eve* 表达的宽阔的条带中转录呢？一个宽阔的条带怎样才能变成细窄的区带呢？解答这个问题有两种模型：

（1）组合模型假设：不同的基因以重叠的窄带表达。每种配对规则基因控制一种窄带发育的形式。不同配对规则基因的窄带相互重叠，这是由于它们各自都越过

了自己的范围而相互重叠。作为这种方式的结果，在细胞囊胚中不同细胞表现出配对－规则基因的不同组合。每个间隔都由自己表达基因的特殊组合来限定，这种组合决定了细胞下一阶段发育的反应。在另一方面，分节基因是由配对规则基因以配对基因受裂缺基因控制的同样方式来调节。

（2）分界模型假设：间隔是由表达的窄带式样限定的。但在界线上与细胞间通讯有关的相互作用导致分隔再进一步细分。*engrailed* 的表达人们假设是由带有和不带有 *ftz*（或 *eve*）的细胞一道触动，这对于触发 *engrailed* 的表达是必要的。

14个体节中的每一个通过体节极性基因的功能进一步地再细分。这种细分在每个体节中都重复进行。例如 Engrailed 区分 A 和 B 分隔。Engrailed 是一种转录因子，但另

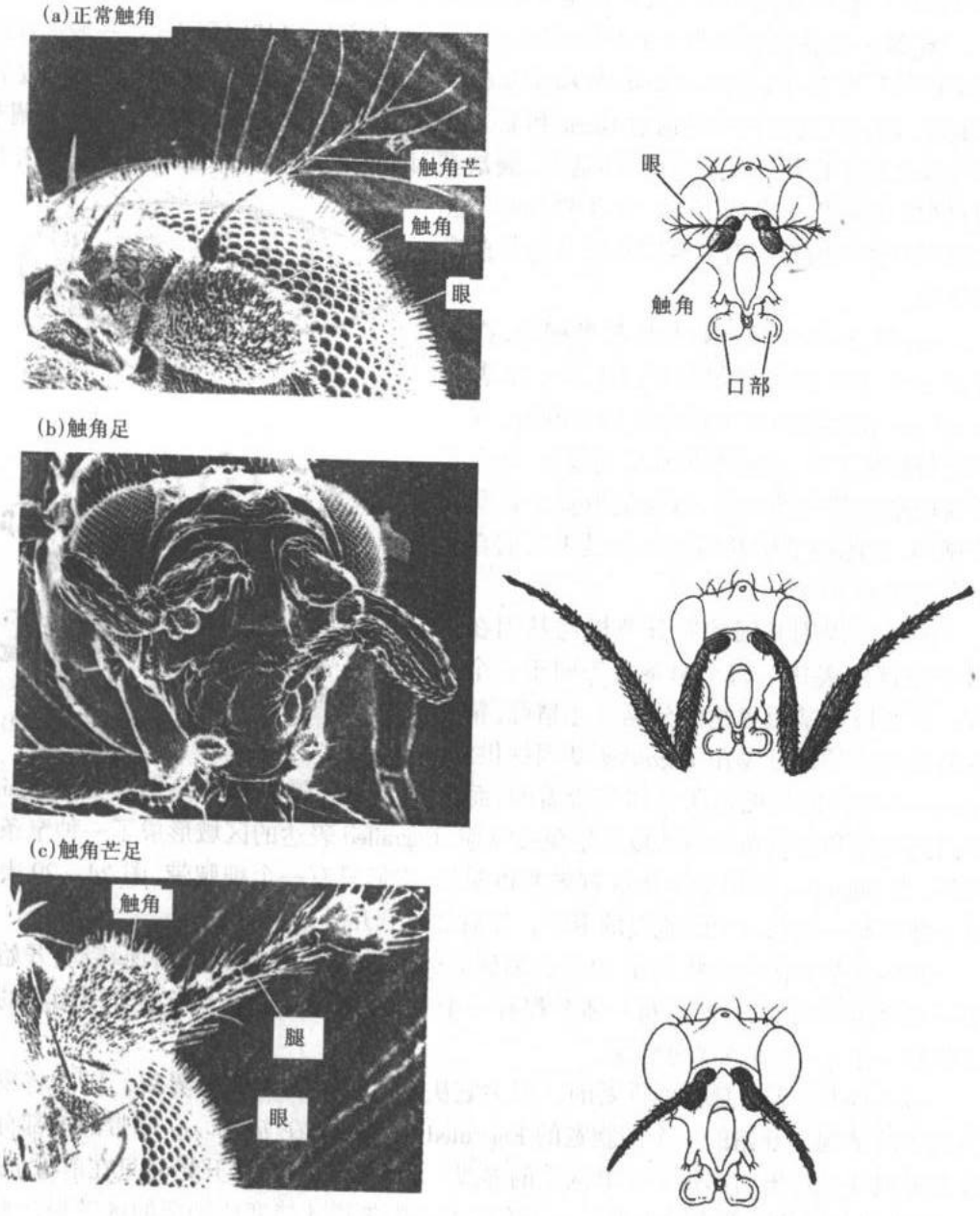


图 24－30 正常果蝇和触角足突变体果蝇的表型。(a)表明触角区的电镜扫描照片和图；(b)果蝇同源异形突变体触角区的电镜扫描照片和图。此突变体在长触角的地方长出了足。(c)果蝇源异形突变体触角芒区的电镜扫描照片和图。(引自 Russell, P.J.: Genetics, 3rd ed., 1992)

一种体节极性基因具有不同的功能。*wingless* 编码一种被邻近细胞所接受的分泌蛋白，这样就能发挥细胞间的作用。*wingless* 是在 *engrailed* 得到表达的细胞前面紧相邻接的一排细胞中表达；这样 *wingless* 归结为识别前面副节的后部界线。*engrailed* 对 *ftz* 和 *eve* 作出反应开始表达，但很快就被从邻接细胞中分泌的 Wingless 蛋白的反馈调节所取代。*engrailed* 的表达对于 *wingless* 的表达是必须的。这样就形成了分界线的样子。现在还不清楚其他体节极性基因怎样确定体节和部分体节的，但是它们的产物，包括分泌蛋白，跨膜蛋白，激酶、细胞骨架蛋以及转录因子，表明在细胞与细胞间的通讯阶段是重要的。

第六节　同源异型复合基因座

一、同源异型基因

体节极性基因控制每一体节的式样，包括极性。同源异型基因决定各体节的特点。同源异型突变使一个体节的细胞发育成另一些体节细胞的形态。有的同源异型突变体一个体节的一部分或整个的体节变成了另一种体节。大部分同源异型基因以

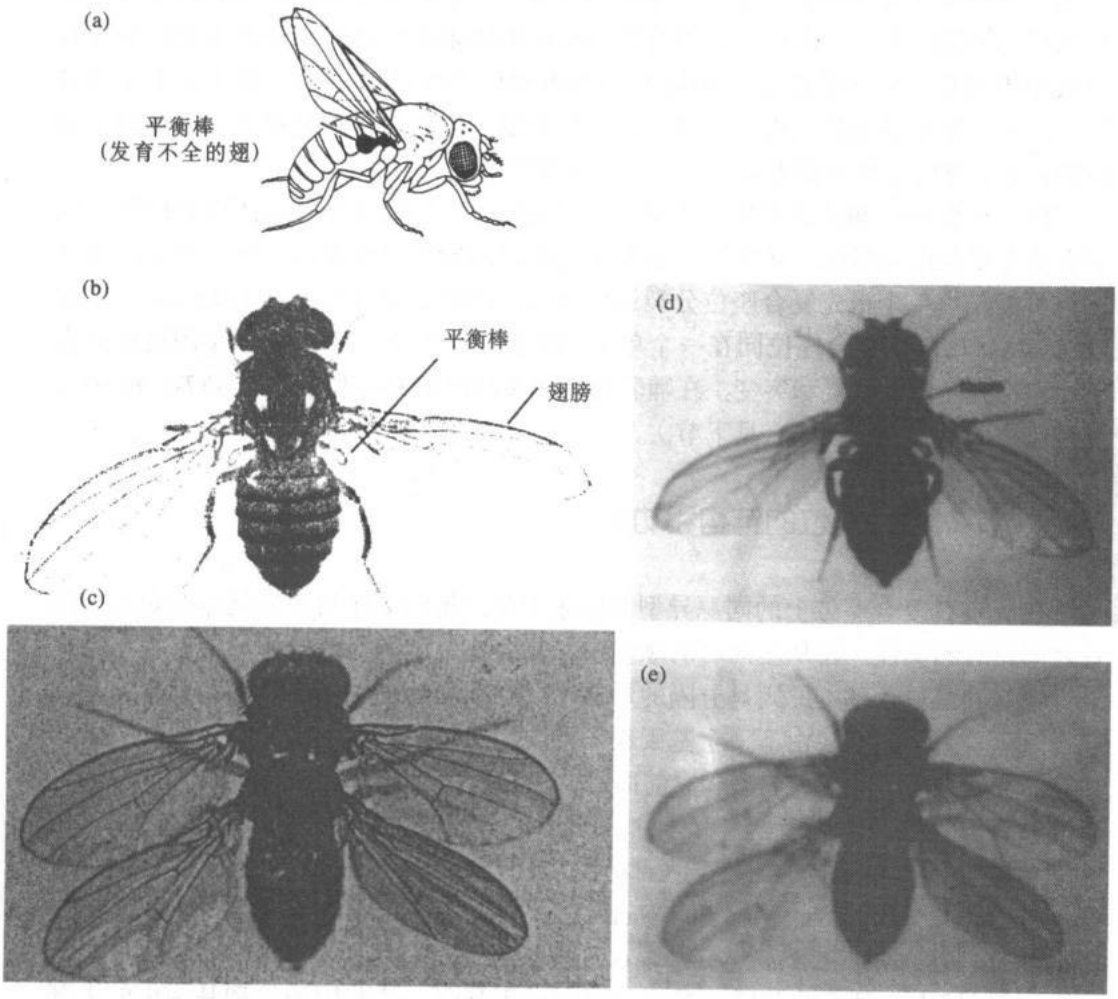


图 24－31　正常果蝇和双胸突变体果蝇的表型。(a)正常果蝇。T＝胸节，A＝腹节，平衡棒位于 T3。(b)正常果蝇(只有一对翅膀)的照片；(c) *bx*，*abx* 和 *pbx* 三对等位基因突变纯合体的照片。(引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 17－32)

和副节相应的方式分别表达。

同源异型基因互以十分复杂的座位间方式进行。很多同源异型基因编码转录因子，它们作用于其他同源异型基因上，就像作用于靶位点一样。一个同源异型基因发生突变，结果影响了另一个同源异型基因的表达。结论是一个突变体的出现不仅取决于一个同源异型基因功能的丧失，而且还取决于其他同源异型基因在对这种功能丧失所作出的反应如何改变其空间模式。

在胚胎形成时同源异型基因发挥其功能。它们的表达依赖于分节基因事先的表达；我们可以将同源异型基因看作为通过分节基因来建立信号的整合方式。同源异型突变导致一个腹部的体节发育成另一种体节。如长触角的地方发育成腿（图 24 - 30）或者在长平衡棒（ baltere ）的地方长出了翅膀（图 24 - 31）。注意同源异型基因并不重新产生一种模式，而仅修饰由一些基因如体节极性基因所决定的细胞的命运。这种修饰是通过控制在特殊部位发挥功能的整套基因的打开与关闭进行的。实际上体节极性基因是在同源异型基因表达达到高峰时起作用的。

有些同源异型突变的遗传性状是罕见的，并使**复合座位**（ complex loci ）得以鉴别。若是一个较大的基因，突变可以将一系列外显子进行作图。复合座位的标志除了间插了一些其他基因以外，通常很大（ > 100kb ），而且存在十分复杂的互补方式，有的具有成对的互补组合，有的却没有。单个的突变可以使不同的复合座位产生形态学表型效应。现已知它们所控制的这些转录单位由于很多调节元件的存在而延伸得很长。对一些仅在一种特异细胞中起作用的启动子和增强子进了互补分析，结果发现很多意外的情况。现在已认识到复合座位元件可控制胚胎不同部分基因的表达，除此之外并没有任何遗传结构的新特点。

两个复合座位和昆虫成体发育有关。*ANT-C* 和 *BX-C* 两个复合座位共同控果蝇所有分节单位的的特征。每个复合座位含有不同的同源异型基因。两个分离的复合座位可能是由单个祖先复合座位分裂进化而来。在甲虫拟谷盗属（ *Tribolium* ）中发现 *ANT-C* 和 *BX-C* 复合座位同在一个单个的染色体位置上。一个新的基因可能来自于原来祖先基因的重复和突变。在哺乳动物相关基因中的一些成员与 *ANT-C* 和 *BX-C* 复合座位的基因是相关的（见下节）。

二、*ANT-C* 复合座位的结构和功能

在 *ANT-C* 复合座位上的同源异型基因簇中基因排列顺序与在果蝇体内表达的位置之间具有相关性。在从左向右的顺序中，复合座位中的每个同源异型基因作用在果蝇较后的区域。基本原则是分隔的形成需要在前面分隔中所表达的基因产物，此产物是延着基因簇顺序由下一个基因编码的，而且加一种新的功能。这样失去了此功能的突变产生了分隔的表型和它前面的分隔相对应。具有 DNA - 结合功能域的一套转录因子由各个基因编码。

果蝇前部（1 ~ 4 副节）的特征大部分是由复合座位（*ANT-C*）来决定的。*ANT-C* 含有一系列同源异型基因，包括 *lab*（labial），*pb*（proboscipedia），*Dfd*（deformed），*Scr*（sex combs reduced）和 *Antp*（antennapedia）。这些同源异型基因位于一个基因簇内约 350kb 的区域中，另一些非同源异型基因间插于其间。这些同源异型基因中的大部分是在不同的发育阶段起到调节的功能。

图 24 - 32 表明作用于身体前部的 *ANT-C*（触角足复合体）复合座位的结构组

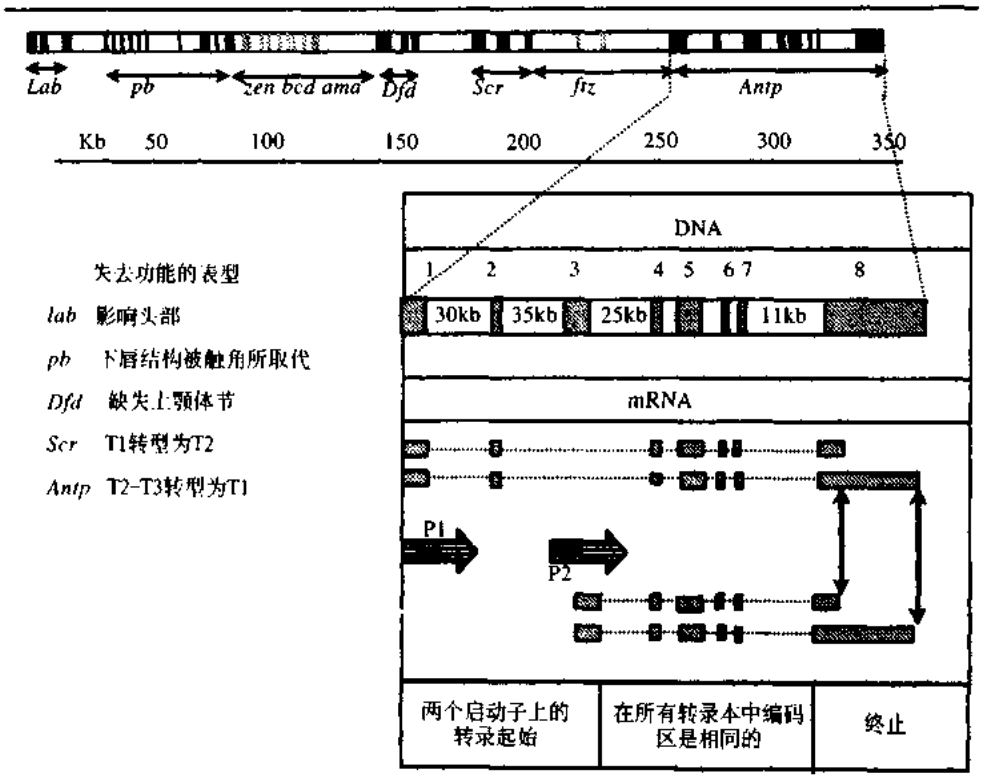


图 24 - 32 *ANT - C* 复合位点的同源异型基因决定果蝇最前面的体节特征。这些基因大小不同，并被其他的基因间插于其间。*Antp* 基因很大且有选择性表达。（仿 B. Lewin: Genes VI, 1997, Fig. 38 - 24）

成。这些相邻接的基因在胚的前部相继表达，从最前的 *lab*(最先起作用影响到头部结构)到最右边的 *Antp* 基因(最后起作用，影响到胸节 T2 ~ T3)。*Antp*（触角足 antennapedia）基因取名就很复杂，此基因的突变使触角变成了第二对腿(图 24 - 30)，或是第二、第三对腿变成第一对腿。*Antp* 的功能通常是控制胸部，它可以促进 T2 ~ T3 胸节以及头部结构的形成。这个基因的功能丧失可导致 T2 ~ T3 胸节发育成 T1 胸节，而头部的过度表达又可导致前部区域发育成胸部的结构。

Antp 有 8 个外显子被很大的内含子所分隔开，总长度为 103 kb 左右，只有在外显子 5 中具有单独的一个开放读框，产生的蛋白为 43 kDa。座位的长度和蛋白质大小之间的差异很大，仅有 1 % 的 DNA 编码了蛋白。

此基因有两个启动子，相距约 70 kb，转录可用两个启动子中的任一个开始启动。一个启动子位于外显子 1 的上游，另一个在外显子 3 的上游。第一个启动子的使用时外显子 3 被删除。从第二个启动子开始的转录本在第 8 个外显子的中部或末端终止。所有的转录本都编码相同的蛋白。这两个启动子的使用经鉴定没有差别，表明其意义可能在于 mRNA 的非转录前导序列的结构不同。

ANT-C 复合座位的其他基因在头部和第一胸节表达。在最前面的间隔中，*lab*, *pb*, *Dfd* 具有独特的表达方式，因此，部分区域的缺失可能由功能丧失突变所致。作用的左 - 右/前 - 后顺序的一个例外情况是 *Scr* 的缺失使有功能的 *Antp* 基因转化的方向与通常的相反。

三、*BX-C* 复合座位的结构和功能

典型的复合同源异型座位是 *BX-C*，即双胸复合体（ bithorax complex ），它含有几组控制胸节发育的同源异型基因，若发生突变会导致腹部形态的改变。当整个复合座位缺失时昆虫在胚胎发育的晚期死亡。复合座位中发生突变的胚胎可以成活，但某些体节的形态会发生改变。图 24 - 31 中所示的转变是一个极端的例子，此例是三重突变，T3A 本来带有平衡棒，突变后变成带有双翅的 T2 胸带，使平常只有一对翅的果蝇长出了两对翅。

BX-C 的遗传图和身体的结构（ T2 ~ A8 ）相关（图 24 - 33 ）。*BX - C* 复合

突变	表型突变	突变	表型变化
abx(anferobitnorax)前双胸	T2/T3A 变成 T2	iab2(infraabdominal)下股部	A2A 变为 A1
bx(bithorax)双胸	T3A 变成 T2A	iab3(infraabdominal)下股部	A3 - A6 变为 A2
bxd(bithoraxoid)双胸状	A1A 变成 T3A	iab4(infraabdominal)下股部	A4 变为 A3
pbx(postbithorax)后双胸	T3P 变成 T2P	iab5(infraabdominal)下股部	A5 变为 A4
Ubx(Ultrabithorax)超双胸	abx,bx,bxd,pbx 总和	iab6(infraabdominal)下股部	A6 变为 A5
		iab7(infraabdominal)下股部	A5 - A8 变为 A3 - A4
		iab8(infraabdominal)下股部	A8 变为 A7

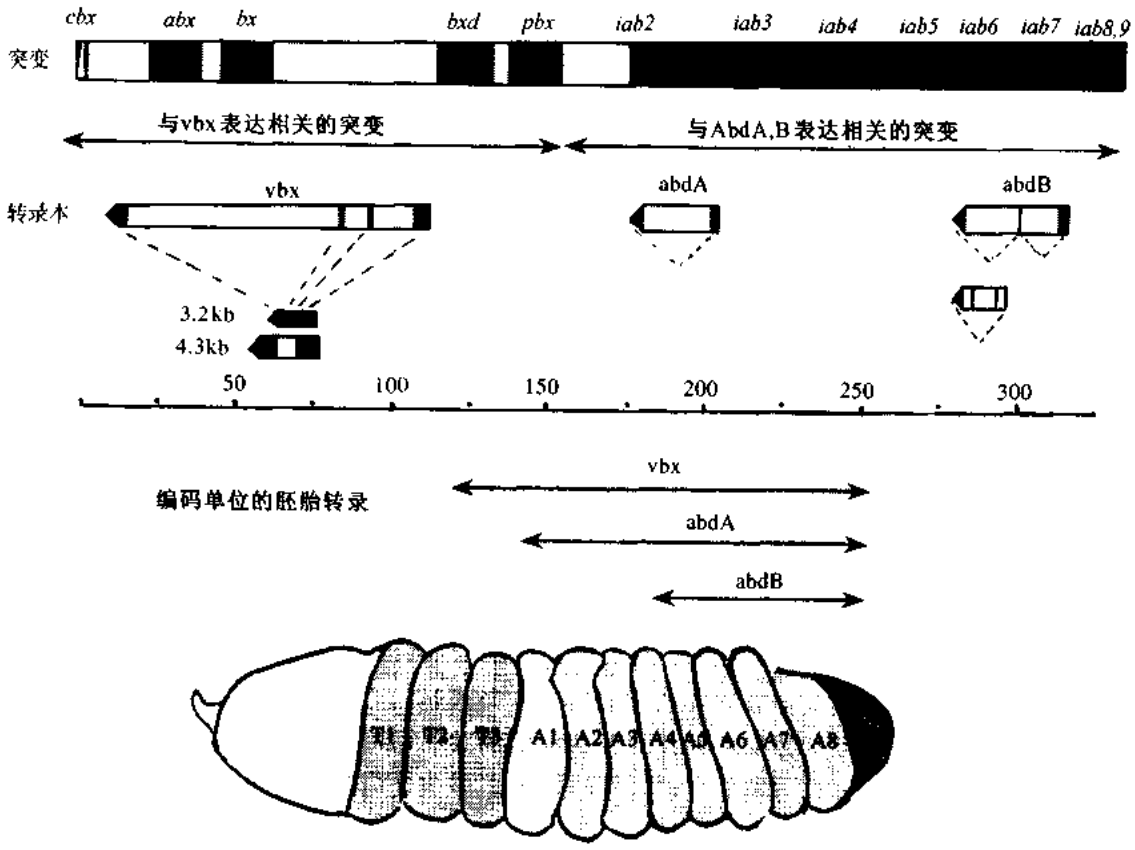


图 24 - 33　双胸复合体座位有 3 个编码单位，各种突变会相继地影响到果蝇体节的发育。发生突变的地方显示出表型的缺失、插入和易位。（参考 B.Lewin: Genes VI, 1997, Fig. 38 - 26）

体影响到果蝇身体大部分体节的发育。和 *ANT-C* 相似， *BX-C* 的重要特点是在复合座位中以一定序列排列的基因其作用的体节也相应地按序排列。遗传图上从左向右的排列的基因所控制的体节从前向后依次排列。

ANT-C 和 *BX-C* 之间的不同之处是 *ANT-C* 的作用明显，其蛋白编码座位是分散的，而 *BX-C* 除了在蛋白编区内的突变效应以外，还显示了以顺式－作用的复合式样。 *BX-C* 长 350kb，其中只有 1.4 %编码蛋白。各种突变可归纳为两类：

(1)3 个转录单位(*Ubx*，*abdA*，*abdB*)产生的 mRNA 编码蛋白。转录单位很大(*Ubx* > 75kb，*abdA* 和 *addB* 都大于 20kb)。每一个都含有几个大的内含子。(*bxd* 和 *iab4* 区域产生的 RNAs 不编码蛋白;其次转录单位是很大的,RNA 产物受到剪接。它们的功能尚不清楚。)

(2)在整个基因簇的间隔中存在顺式作用突变。它们控制了转录单位的表达。任何特殊类型的顺式作用突变都可能发生在一个很大的区域中。图中表示的位置仅仅是近似的。在交界处也可发生各种突变。

此复合座位以前用两个功能区来表示。在 *Ubx* 功能区中的突变首先被描叙，它们以胸节 T2P ~ T3P 和腹节 A1A 作为它们 的靶(此和 5 ~ 6 副节相对应)。这些突变或位于 *Ubx* 转录单位中,或位于控制 *Ubx* 的顺式作位点。在 *Ubx* 功能区中的突变根据其表型命名。*bx* 和 *bxd* 型是根据各种突变而鉴定出的,每种跨越 10kb 左右，*abx* 和 *pbx* 突变由缺失而产生,范围分别为 1 ~ 10kb 不等。

iab 功能区的突变发现较晚,它们以腹节 A1P ~ A8P(副节 7 ~ 14)作为靶子。这些突变或位于 *abdA*, *B* 转录单位,或在控制它们的顺式作用位点。在 *iab* 功能区中顺式作用突变被系统地命名为 *iab* 2 ~ 9。这些突变影响到单个的间隔,有时也影响到间隔的邻接位点。

延着基因簇从左向右转录本在果蝇胚胎后部有所增加。如图 24－33 下部所示，出现了重叠的式样。 *Ubx* 在 *T2P* （副节 5 ）中的表达有一个前部的分界线。 *abdA* 的表达从 *A1P* （副节 7 ）开始，而 *abdB* 在后部的 *A4P* （副节 10 ）中开始表达。

Ubx 的转录研究得最为深入。 *Ubx* 转录单位长约 75 kb，由于不同的选择性剪接产生了几种短的 RNAs 。 4.7 kb RNA 短暂地出现，然后被 3.2kb 和 4.3kb 的 RNAs 所取代。后两种 RNA 的共同特点是都含有初始转录本两端的序列。当然其他尚未鉴定的 RNAs 也可能有同样的特点。

这两种 RNAs 之间的差别究竟有何意义现在尚不能确定。最前和最后的外显子长度相等，中间的外显子比较小，来自长转录单位中的相同外显子可能参与通过选择性剪接而产生 RNA ，但现在尚不知道通过各种表达方式产生的 *Ubx* 蛋白有任何功能性的差别。

Ubx 转录本的一个表达模型假定选择性剪接产生 RNAs （包括但不限于 3.2 和 4.3kb 的 RNA ）具有编码区的开始和末端。但中部的外显子不同。根据此前提，由此转录单位编码的整套蛋白用第一个共同外显子的抗体进行了直接定位。 Ubx 蛋白仅由 *Ubx* 区域编码。

发现 Ubx 蛋白所在的体节与转录的位置相一致，即在 T2P ~ A1 中。在 A2 ~ A8 中也有低水平的存在。这表明 *Ubx* 单位编码整套的相关蛋白。 *Ubx* 结构区的突变会影响到体节中 Ubx 蛋白浓度。 Ubx 定位在核中，作为一般的转录调节因子，它们具有同源异型结构的 DNA －结合区。

通过功能丧失突变能帮助我们了解 BX-C 复合座位的一般功能。若缺失了整个的

复合座位，幼虫的体节是不能发育的。若以副节（可能是作用单位）来表示的话，所有的副节都以第 4 副节同样的方式产生差异；果蝇胚沿着它的体长有 10 个重复的 T1P/ T2A 的结构。代替了 5 到 14 副节之间的间室。其实 *BX-C* 功能的缺失使 *Antp* 的表达贯穿腹节，这样使这些体节具有了 *Antp* 所决定的一些特征；BX-C 功能对于加一些后部的信息来说是必要的。

每个转录单位按照其表达方式作用着连续的体节。若仅存在 *Ubx* 的幼虫只有副节 4（T1P/T2A），副节 5（T2P/T3A）和 8 个重复的副节 6（T3P/A1A）。此表明 *Ubx* 的表达对于 A1A 前的间室来说是必要的；*Ubx* 也在较后面的体节中表达，但在野生型中 *abdA* 和 *abdB* 也存在；若它们被除去的话，*Ubx* 在所有后部体节中所起的作用和常在第 6 个副节（T3P/A1A）中所起的作用相同。

Ubx 加上 *abdA* 使副节 7、8 和 9 增加野生型的模式。另外 *Ubx* 加 *abdA*，能影响到 A3P/ A4A。缺失 *abdB* 时使得较后部的间室外缺失。表明 *AbdB* 对于副节 10 ~ 14 的发育是必要的。

ANT-C 和 *BX-C* 复合座位功能的一般模型认为其附加功能是向后面的方向加上合适的连续的体节。其功能的实现依赖于重组方式，在重组中增加了连续的基因产物所赋予的新的特异性。这就解释了 *ANT-C/ BX-C* 复合座位中一个基因发生功能丧失突变常由排列在突变基因前的一些基因决定表型。

Ubx 在较前一些的体节中表达具有与丢失的功能相反的作用。这些体节通过发育出现前部的表型。当 *Ubx* 放在头部表达时，前部的体节变为副节 6 的表型。这样若无 *Ubx* 的表达就会产生同源异型转化，转化中后部体节获得了前部的表型。*Ubx* 的过量表达也会导致同源异型转化，在转化中前部体节获得了后部的表型。

这种关系是作为整个的一个基因簇，并解释了顺式作用突变的特点。这些调节的突变使蛋白质丢失了部分结构域或增加新的结构域。这样使得表型变化或丧失功能，或增加了功能，或两者兼之。在单个的间室中最常见的是丢失了功能。例如 *bx* 特异控制 *Ubx* 在间室 T3A 中的表达；一旦 *bx* 发生突变 *Ubx* 在 T3A 中就不表达，使 T3A 转变成较前一些的间室 T2A。这个例子是在复合座位中个别顺式作用突变的典型例子。每个转变的靶间室发育成其前面的相应体节。顺式作用位点突变在染色体上排列的序列反应了它们所作用的间室的排列序列。因此 *Ubx* 在副节 4、5、6 中的表达相继受到 *abx*（作用于副节 5）和 *bx*（作用 T3A）等的控制。

在 *BX - C* 复合座位中只有 3 个基因，有 2 个主要的问题：首先是这三种蛋白怎样组合才能使 10 个副节各具特征？一种可能是在不同的区域中存在数量差异，使得在靶基因中以同样的方式作出不同的反应，就是前面已讨论过的分节基因的组合功能。第二个问题是这些蛋白在不同的组织中如何发挥它们的功能？前面已介绍过产生表型的方式；其他组织的发育的控制是以平行的方式，但并不相同。例如 *Ubx* 是在所有后部体节直到 A8 的表皮和中胚层中表达，在体节 A7 的后面被抑制。后部的分界线反应受到 *abdA* 的抑制。当 *abdA* 发生突变，*Ubx* 的表达可以扩展到后部的中胚层中。

为什么调节发育的基因和编码一般蛋白的基因不同。它们巨大的长度是否是选择性地产生产物所必须呢？它是否可能与时间选择机制、决定转录单位的长度有关呢？在典型转录速率上，转录 *Antp* 可能要花 100 分钟左右。这对于果蝇胚胎发生 22 小时这段时间而言占了很大的比例。

沿着胚胎从前到后的过程，*ANT - C* 和 *BX - C* 座位的一些基因表达模式发生变

化，是什么在控制它们的转录呢?在分节座位中，同源异型座位部分受到发育中前期基因的控制，部分受到它们本身的相互作用的控制。例如 *Ubx* 的表达因 *bicoid*、*hunchback* 或 *kruppel* 的突变而改变。和副节边界有关的前部表达分界线由 *ftz* 和 *eve* 来限定。一般的原则是所有这些基因的功能都受到转录控制的调节。要么被激活，要么被抑制，而且这些基因产物可通过质量和数量两者组合来发挥特殊的作用。

四、 同源异型盒

同源异型盒（ homeobox ）是同源异型基因中都具有的保守基序。

控制果蝇发育的 3 组基因：母体基因，分节基因和同源异型基因。其中的一组调节另外两个中的一组基因和编码结构的靶基因。调节基因之间的相互作用是通过分析其他基因发生突变时相应基因的表达改变来确定的。但目前只鉴别出少数几个结构靶基因，3 组调节基因作用在靶上，使得身体的各部分产生分化。

分节基因编码的蛋白调节转录，3 个裂缺座位（ *hb*，*kr*，*kni* ）基因含有锌指模体、在转录因子 TF 和 SP1 中的模体负责和 DNA 连接。裂缺基因中其他座位的产物也有 DNA －结合模体；*giant* 编码的蛋白具有氨基酸拉链模体；而 tailless 编码的蛋白与类固醇受体相似。这表明裂缺基因的一般功能是转录调节。

在很多同源异型基因和分节基因中发现了保守的模体。普遍的模体是同源异型盒，它位于各种分节基因和同源异型座位转达录单位 3′端附近的 180bp 的区域。此序列是开放读框的一部分。在这个序列中密码子的第三个碱基常发生改变。同源异型盒的大小及序列都是固定不变的。约有 40 个基因含有同源异型盒，而且这些基因几乎都和发育调节有关（同源异型盒是通过它在同源异型基因中**超优势度**（ predominance ）而首先被鉴别出来的，因此得名）。由同源异型盒编码的蛋白序列称之为**同源结构域**（ homrodomain ）。

果蝇的同源结构域分成为几组，主要是由 *BX-C*/ *ANT-C* 复合座位中同源异型基因组成。它们被称为触角足组。它们的同源结构域具有 70 % ~ 80 %的保守性，通常是存在于蛋白质的 C 端。在相关的 *engrailed* 和 *invected* 基因中发现了特殊序列，它和触角足组只有 45 %的同源性。其他类型的同源异型结构区序列存在于 2 ~ 4 个基因中。

在果蝇中很多含有同源异型盒的基因组成了基因簇。在 *BX-C* 基因簇中有 3 个同源异型基因具有同源异型盒。在 *ANT-C* 复合座位中含有一组 5 个带有同源异型盒的同源异型基因，和 4 个也含有同源异型盒的其他基因。在 *BX-C* 和 *ANT-C* 中的同源异型基因有时也用 HOM-C（ homeotic-complex ）基因来表示。

在前－后轴上 *HOM-C* 的功能是什么呢？据推测同源异型结构区上带有不同的氨基酸序列可以识别 DNA 上不同的靶序列。在不同的蛋白之间交换它们的同源异型结构区，结果表明这些蛋白主要的特异部分就在于同源异型结构区。但它们与特异的 DNA 靶位点的结合能力并不能表明它们的特点。例如，这些蛋白中的某一些在前后呼应中要么能激活转录，要么能抑制转录，它们的活性依赖于另一套能结合 DNA，而不识别 DNA 的蛋白。在这一组基因中密切相关的一些成员的同源结构区之间的相似性表明它们能识别靶位点的重叠模式。这样为组合效应开辟了一条途径，使带有相关同源异型结构域的蛋白之间就能像依赖质的差异一样依赖于量的差异，对同一结合位点，进行竞争。在有的情况下，不同的同源异型蛋白识别 DNA 上相同的靶位

点，这也提出了有关它作用特异性的难题；估计 DNA 结合位点上仍存在着微小的差异或者存在着相互作用，例如蛋白质之间的相互作用。

同源异型盒基序在进化中广泛存在。在蛙，小鼠和人的 DNA 都发现同源异型盒的存在。在蛙和哺乳动物基因在胚胎发生的早期表达增强了与果蝇一些基因的平行关系，表明可能这些含有同源异型的基因与各种生物胚胎发育中基因调节有关。

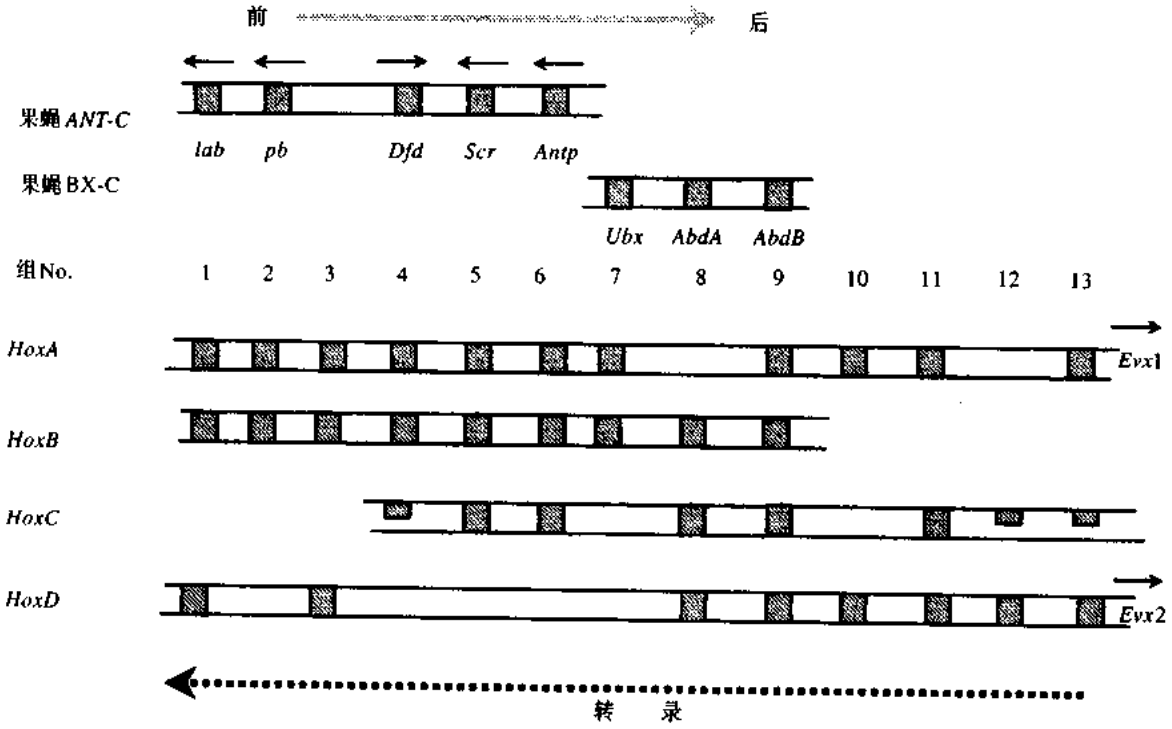


图 24 - 34　小鼠和人的基因组中有 4 个含同源异型盒的基因簇，这些基因的顺序反应了在此区域中它们的表达是沿着前 - 后轴进行的。*Hox* 基因的排列与果蝇同源基因是一致的，特别是第 1, 2, 4 和 9 组更为明显。这些基因的命名是根据“组”和“簇”，如 *HoxA1* 就是 *HoxA* 组中最前面的一个基因。除了小鼠中丢失了 C 基因簇中的某些基因（用半个方框表示）以外，*Hox* 基因都存在于人和小鼠的基因组中。（仿 B. Lewin: Genes VI, 1997, Fig38 - 27）

在哺乳动物（可能所有动物）中那些与 *HOM-C* 组相关的基因都具有一个显著的特点：它们像 *BX-C*/ *ANT-C* 复合座位一样组成基因簇。其中单个的哺乳动物的基因称为 *Hox* 基因，其基因簇可达 20 ~ 100kb，含有的基因可多达 10 个。4 个 *Hox* 家族中的基因含有同源异型盒，它们在小鼠和人类的基因组中被描叙。图 24 - 34 表示这四个基因簇与果蝇中两大基因簇的比较。

通过比较同源异型盒的序列表明哺乳动物中 *Hox* 基因的排列和果蝇的同源基因是一致的。如 *HoxA4* 和 *HoxB4* 与 *Dfd* 相关。当哺乳动物的 *Hox* 基因簇和果蝇的同源异型盒基因簇的关系确定之后，人们就能识别一系列与 *ANT-C* 和 *BX-C* 基因簇相关的基因。哺乳动物 *Hox* 基因 1 ~ 9 组已被确定和 *ANT-C* 和 *BX-C* 座位的序列相一致。10 ~ 13 组是通过串联重复产生的，与 9 组产生差异（图 24 - 35）。在每一个基因簇中的相应座位有时称为平行进化同源基因（paralogs），例如 *HoxA4* 和 *HoxB4* 就是平行进化同源基因。若果蝇和哺乳动物某一点上产生分歧的话，那么果蝇就可能存在只有单个的复合座位，（如 *ANT-C* 和 *BX-C* 不同源），而复合座位含有的各基因都限定了前后极性。文昌鱼（*Amphioxus*）进化路线是和脊椎动物平行的。这种生物有单个的 *Hox* 基因簇，含有平行进化同源基因组中的一个成员；是原始基因簇的典

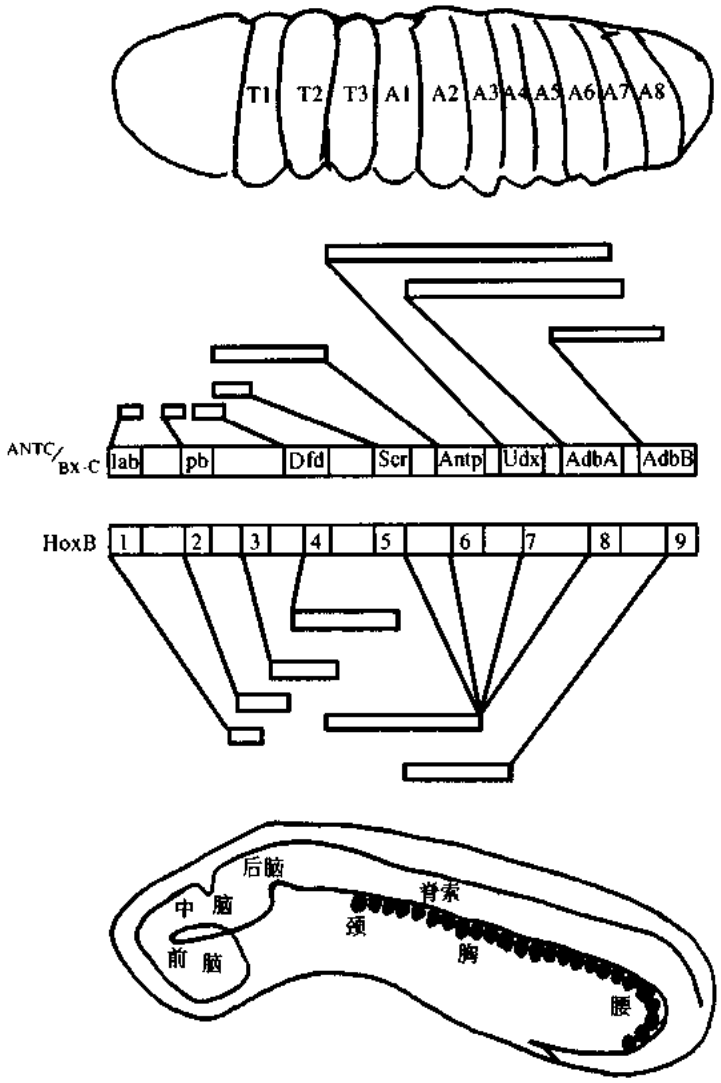


图 24 - 35 *ANT-C/ BX-C* 和 *Hox B* 表达模式的比较表明各个基因的产物在胚中的定位与在基因簇中的排列(从左到右)是一致的。表达模式显示在果蝇表皮中表达的区域(10 小时)和在小鼠(12 天)胚的中枢神经系统中的表达方式。

型。在进化中果蝇的这些基因分成为两个基因簇，而哺乳动物整个一组基因发生了重复，在重复之后个别成员从各个复合座位中丢失了。

这些基因在小鼠和果蝇中的排列和空间表达的方式都表现出平行性。在 *Hox* 基因簇中的基因在胚胎中表达方式与两者在基因组中的组成对应。这样产生了一种很大的可能性，即基因簇不仅分担了共同的进化，而且在基因组的组成与果蝇，小鼠空间表达的相关性中维持了相同的功能，在同源基因之间还存在着某些一致性。小鼠的 *Hox D4* 或 *HoxB6* 在果蝇中的异位表达产生的同源异型转化与 *Dfd* 或 *Antp* 的同源异型表达的结果完全相同，从而证实对二者关系的推论。由于小鼠和果蝇蛋白之间的同源性，结果使它们保持了特有的异型结构域，这就加强了这样的观点，即结构域决定特异性。

Hox 基因很小，但具有大量的蛋白编码单位。小鼠的 *HoxB* 基因簇长 120kb 左右，含 9 个基因。这些基因在基因组中的位置和胚胎中表达两者之间的连接与果蝇中的情况相类似，但前面只描述了基因的本身，现在尚没有掌握有关顺式作用位点的信

息。这是因为在脊椎动物中通过各种突变来得出结论是十分困难的。但已发现 *Hox* 基因的控制仅通过上游区域的启动子和增强子。这还得看是否广泛地存在着突变基因的副本，是否存在果蝇同源异型基因的复合调控区。

在果蝇中每个基因都是唯一的，但在脊椎动物中基因簇的重复使多基因（平行进化同源基因）具有相同或十分相似的表达模式。若平行进化同源基因有冗余或部分的冗余功能，那么缺乏一种产物至少可以被另一个基因簇相应基因产物所代替，因此突变的效应也就可减少到最小的程度。

小鼠中的 *Hox* 基因通过重排也可被破坏而导致隐性致死。在头部和胸部的结构变异体中 *HoxA1* 和 *HoxA* 缺失。不是在所有结构突变体中基因都错误表达，表明确实存在功能的冗余性，即 1 组或 3 组的 *Hox* 基因能弥补某些组织 *HoxA* 基因缺乏，但不能弥补所有其他组织对 *HoxA* 基因的缺乏。

在小鼠中同样的突变能引起同源异型转化，但要比果蝇中少。例如 *HoxC8* 的缺失使某些骨节出现前部骨节的表型。

Hox 基因的异位表达已成功地用于论证，这些基因能使一些体节的特征发生相应的转化。最常见的是果蝇的前部体节出现了后部体节的表型。其实同源异型基因的表达是加上了附加信息。这些附加信息控制了后部体节的特征。在小鼠中也获得了相似的效应。但模式不完全一致。

这些清楚表明 *Hox* 基因与果蝇决定前后轴发育模式的基因簇很相似。可能存在着 *Hox* 基因表达的组合密码子或者可能在平行进化同源基因之间功能冗余的程度不同。但目前尚未能提出它们在决定方式中所引起作用的模型。

Hox 基因座位结构仍难以解释为什么基因簇基因组的位置与胚胎的表达相一致？在进化中是如何维持下来的？一种解释是基因表达的全面控制使得一个基因只有在基因簇中才能适当地表达。但这并不真实，至少在个别情况中一些基因可从基因簇中取消掉。分析了启动子区，表明 *Hox* 基因可受控于一系列的启动子或增强子元件，启动子和增强子控制了整个的表达模式。通常这些元件是在转录起始点的上游区域。例如 *HoxB4* 能在含有这些元件的质粒中表达，并可以制备转基因鼠，但为什么在进化压力下使一些基因排成有序的基因簇。一种可能是一个基因的增强子可能被埋入在另一个基因内。我们可以将一个基因易位别处时，看它是否有功能来加以判断它是否属于这种情况，若它含有相邻基因的调节元件的话，那么它被取消将影响到其他基因的表达。

第七节　动物的性别决定和分化

在生物界中都存在着两种性别，在高等生物中性别决定和分化的分子机制是多种多样的，我们将以果蝇和哺乳动物为例来介绍二种完全不同的性别分化的机制。

在有些生物中环境对性别的分化有着很强的作用。例如海里的环节动物 *Ophryotrocha*，当它幼小的时候分化成产生精子的雄性，当它长大以后又变成产卵的雌性。若将成年雌体切断，那么一段蠕虫又可变成雄性，表明是其身体的大小在控制性别的表型，而不是年龄。

在大部分高等生物中，性别决定存在着遗传基础。但在某些物种中性别的遗传控制是和环境因子共同起作用。例如在膜翅目昆虫中的蜜蜂、黄蜂，其性别是由性别座位上等位基因来决定的。若任何特异的等位基因无论是纯合还是半合，个体都发育成雄

性;若此座位的两个等位基因是异合的话则发育为雌性。由于性别座位高度的多态性,也就是说存在很多复等位基因,因此受精卵几乎都是异合的,以后将发育为雌性。蜂皇(2n = 32)能产两种不同的卵,由其纳精囊(一种器官接受交配时注入的精子)的括约肌来控制。它产生的一种受精卵,是具有 32 条染色体的二倍体,将来孵化成雌蜂,她还产生另一种未受精卵,只含有 16 条染色体,将来发育成单倍体的雄蜂。二倍体雌性卵能发育成工蜂和蜂皇,此完全取决于发育的条件。

在很多生物中,性别决定是和一对异形染色体有关。XY 性别决定存在多次独立的进化。这样构成 XY 性别决定的分子基础也可能是完全不同。例如果蝇和哺乳动物就是很好的例子。在果蝇和哺乳动物中身体的不同部分存在性别的二形性特征,这些第二性征(secondary sexual characteristics)下面将会集中讨论。

一、果蝇的性别决定

果蝇的性别分化需要很多调控环节,我们先看一下总的途径(图 24 - 36),

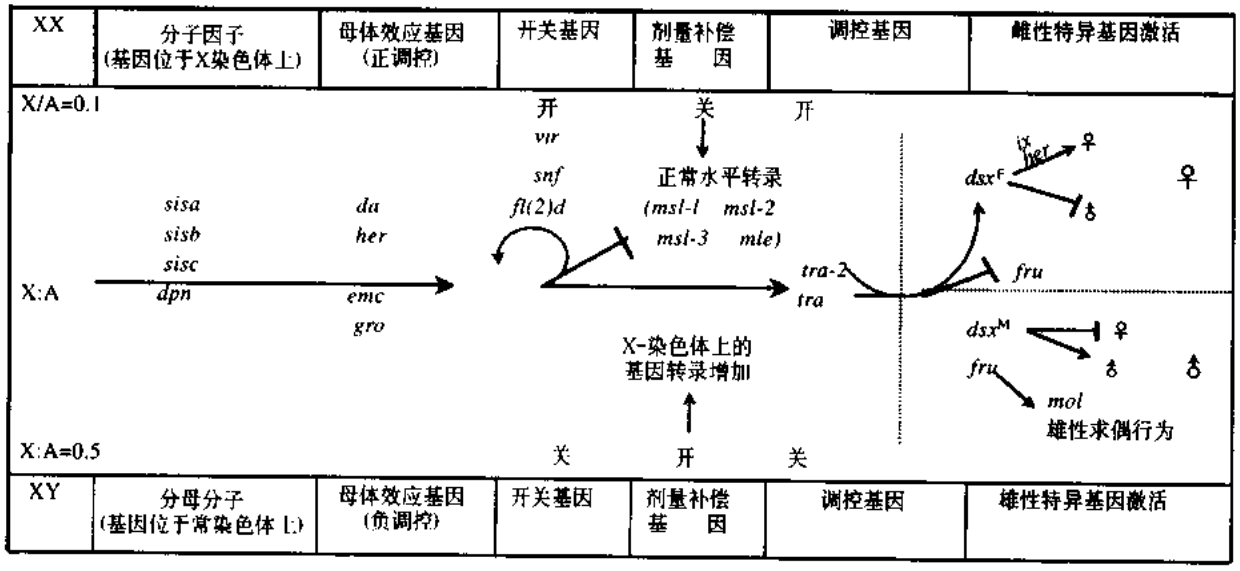


图 24 - 36 果蝇性别决定和分化的途径及调控。

然后再介绍细节。果蝇早期胚胎中性指数(X : A)决定了性别的分化。与性指数相关的性别决定途径是由 Thomas Cline Baker (1984, 1993) 及其同事们建立的。在这些途径中有主调节基因和系列性别特异基因,它们涉及到 RNA 的不同剪接,产生不同的雌性和雄性的转录因子。

在此调节途经的第一步是由 X : A 的比例决定的。性指数使转录因子具有特殊的浓度。而这种特殊的浓度充当了主调节基因的开关。当开关打开时 RNA 能正常剪切而产生下一步调节的活性转录因子,从而激活雌性特异基因,使胚胎发育为雌体;若开关是关闭的,那么因剪接的方式不同选择性地产生了另一种转录因子,它能激活雄性特异基因,使胚胎发育为雄体。

在讨论性别决定调节途径的细节之前,我们先来了解一下性别特异突变。性别转换基因 *tra* (*tra*nsformer) 发生突变使得 X : A 比率为雌性(X : A = 1)的个体发育成雄性表型,但核型并不改变。双重性别基因 *dsx* (doublesex,) 突变使个体变

成为间性，而不受性染色体的控制。还有一种突变会导致一种性别致死，这就是性别致死 *Sxl*(sex-lethal)基因，其突变有些等位基因是隐性的，可特异导致雌性致死；而另一些等位基因是是显性的，可导致雄性致死。无女儿 *da*(daughterless,)基因是母体基因，对性别决定具有母体效应，*da* 纯合时母体不产生“女儿”，而“儿子”发育却正常，这是由于其产物对 *Sxl* 的早期表达起着正调控作用。在卵产生时 *da* 若无活性，那么在受精后无论 X 染色体数量如何，都不能激活 *Sxl*，结果所有的子代都向雄性方向发育，但 XX 个体因 X－连锁基因的剂量比 XY 多了一倍，因剂量不适而死亡。特大刚毛基因 *emc*（extramacrochaetae）也是母体效应基因，其作用与 *da* 相反，是 *Sxl* 的负调控因子，*emc* 的突变将抑制雄性方向的发育。合子中染色体上的 *da* 和 *emc* 两个基因并不影响合子本身的性别发育。

1. 果蝇 Y 染色体的功能

果蝇的 Y 染色体与精子发生有关，而尚未发现它在性别决定中有任何功能。X/O 型的果蝇可以发育为雄体，Y 染色体在体细胞中可以丢弃，但在初级精母细胞中都处于活跃状态，X/O 雄蝇精子发生受到严重的干扰，产生的精子无活动能力。

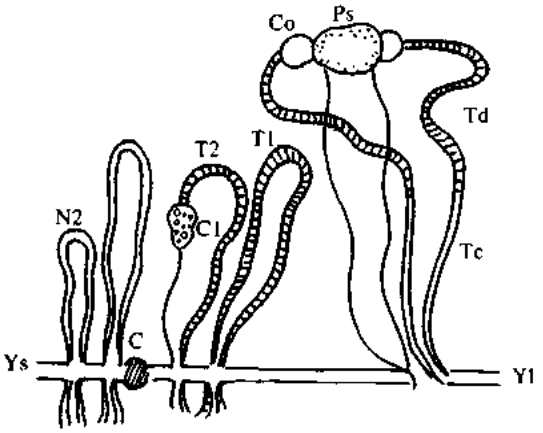


图 24－37　果蝇初级精母细胞双线期的灯刷环。C. 着丝粒；Ns.(noose)：基索；Cl.(club-like)棒状；Co.(Cone－like)球果样结构；Tr(tubular ribbon)，管状带；Th(thread)，丝状。

Y 染色体有 6 个雄性可育性基因（长臂 4 个，短臂 2 个），而在 X 染色体上无相应的等位基因。这 6 个基因只转录，不翻译，每个基因转录成一种灯刷环，这种灯刷环是一个转录单位，它们和一些蛋白质结合形成不同形态的 RNP，分别命名为套索状，棒锤状，管带状，丝状，球

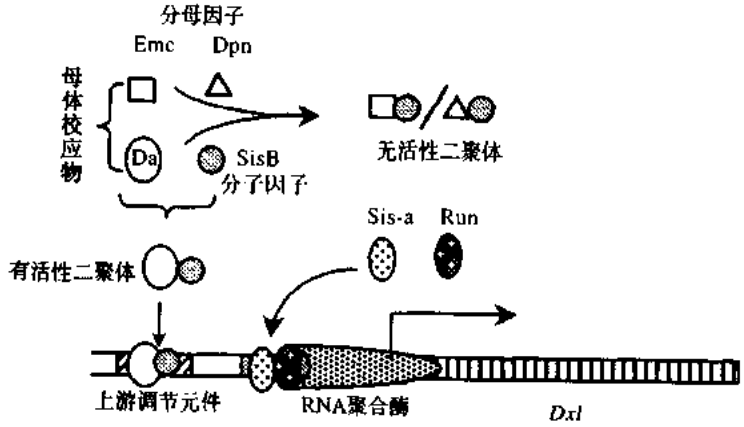


图 24－38　母体效应基因，分子基因和分母基因对 *Sxl* 转录激活的作用。

果样结构(图 24－37)。一个或多个灯刷环失活都会引起雄性不育。可见是精子发生不可缺少的。其特点是①不翻译；②RNA 很大，约为 1500kb，③有 Ys 和 Ya 两种序列。Ys 是特异序列，约有 200～2000 个拷贝；Ya 是 Y－结合序列(Y-associated sequence)，一个家族只出现在一个环上，长约 400b；(4)能形成二级结构；5′能富集蛋白，形成 RNP。

Y 染色体对某些蛋白起着调控作用。X/0 雄果蝇精巢中有 3 种蛋白(包括微管蛋白)的量明显下降，而这些蛋白都不是 Y 染色体编码的。微管蛋白的 mRNA 在 X/0 雄果蝇中并不减少，可见 Y 染色体可能在微管蛋白翻译水平上起到调节的作用。X 染色体上的星状(satellite)位点编码 P17 蛋白，在 X/0 雄蝇的精母细胞中该蛋白由于过多合成而形成结晶，Y 染色体对其有调节功能，调节位点不在灯刷环上，而在可育位点 K12－K13 之间。有部分 Y－RNA 存在核内，在精子细胞中仍存在这些 RNA，它可以与 DNA 互补，通过反义抑制进行调节。Y－RNA 的二级结构是结合蛋白所需，而这些蛋白可能和凝聚与去凝聚的过程有关。从上面的介绍不仅说明了果蝇 Y 染色的功能，而且使人们不难理解了为什么果蝇的性别决定似乎完全和 Y 染色体无关。

2. 性指数的作用

果蝇虽然是 XY 型的动物，但 Y 染色体并不起作用，性别决定取决于性指数。性指数决定性别的机制是建立于卵巢和早期合子中合成的基因产物的相互作用。这些基因分为三类，即母体效应基因，其中 *da*，*her*(hermaphrodite 两性的)为正调控因子，*emc*、*gro*(groucho 喜怒的)为负调控因子。合子中表达的 X 染色体上的基因和常染色体上的基因，它们的产物相互作用，为 *Sxl* 的激活提供信号(图 24－38)。目前已搞清了其中一部分细节。

母体基因产物和部分合子基因产物是 HLH(helix-loop-helix)家族中的转录因子。它们可以是同二聚体，也可以是异二聚体。HLH 蛋白可能受到其他不同二聚体蛋白比例的调控，即分子因素和分母因素。在母体基因和合子基因的相互作用中，X－染色体基因产物的作用称为**分子因素**(numerator elements)，而常染色体基因产物的作用称为**分母因素**(denominator elements)。雌性(XX)的 X－连锁的分子因素是雄性个体(XY)的二倍，而在两性中分母因素的产物量是相同的，这样雌，雄的 X∶A 的比率分别为 1.0 和 0.5。性指数决定性别的机制是将合子中分子因素在两性中的量的差别转变为 *Sxl* 基因转录的“全”或“无”质的效应。

现已发现的分子因素有 *sis-a*(sisterless-a 无姐妹-a)。*sis-b*，*sis-c*，*run*(*runt* 侏儒基因)，分母因素有 *dpn*(deadpan 无表情的)。其中 Sis-a 蛋白含锌指结构；Sis-b、Dpn、Da、Emc 蛋白均可形成 HLH 结构。一种可能是 Sis-b 和其他的分子基因产物或与 Da 蛋白，或相互间，或与常染色体的分母基因以浓度依赖性方式形成 HLH 二聚体。其中某些异二聚体将作为 *Sxl* 基因的激活转录因子，而有的可能以其他方式进行调节。

分子基因编码蛋白产生得越多，那么由它们形成的激活二聚体的水平越高。在雌果蝇(X∶A＝1.0)中有足够的激活二聚体去打开 *Sxl* 基因，而在雄果蝇中(X∶A＝0.5)转录因子的浓度就不足以激活开关基因 *Sxl*。分母基因的产物和起负调控的 Emc 可能与母体效应基因产物，与分子基因产物形成没有功能的二聚体(图 24－38)。

3. *Sxl* 基因是主开关

在果蝇囊胚形成前或形成时，通过对性指数的计量决定是否打开 *Sxl* 基因。*Sxl* 基因是性别决定的总开关，它能永远保持在“开”和“关”的状态。当 X ： A 转录因子浓度达到一定时(X ： A = 1.0)，*Sxl* 基因处在“开”的状态。而在雄体中 X ： A = 0.5 转录因子未能达到一定的域值，*Sxl* 基因处于关闭状态。

Sxl 基因有早期和晚期两个启动子(P_L 和 P_E)和 8 个外显子，在第 2 个外显子中有翻译的起始密码子 AUG，而在第 3 和第 8 个外显子中有终止密码子 UGA（图 24 - 39）。早期启动子是调节型的，由母体效应基因，分子、分母因子共同调控。这个启动子开始的转录是选择性的，即可以跳过第 3 个外显子，这样也就避免了中途的终止密码子，使翻译可以一直达到最后一个外显子中的 UGA，产生有活性的 *Sxl* 蛋白（350aa 组成）。囊胚形成后 X ： A 转录因子和母体效应基因的产物已不复存在，P_L 已不能被激活，而晚期启动子 P_E 是组成型表 0.0000000 达的，无需母体基因及分子、分

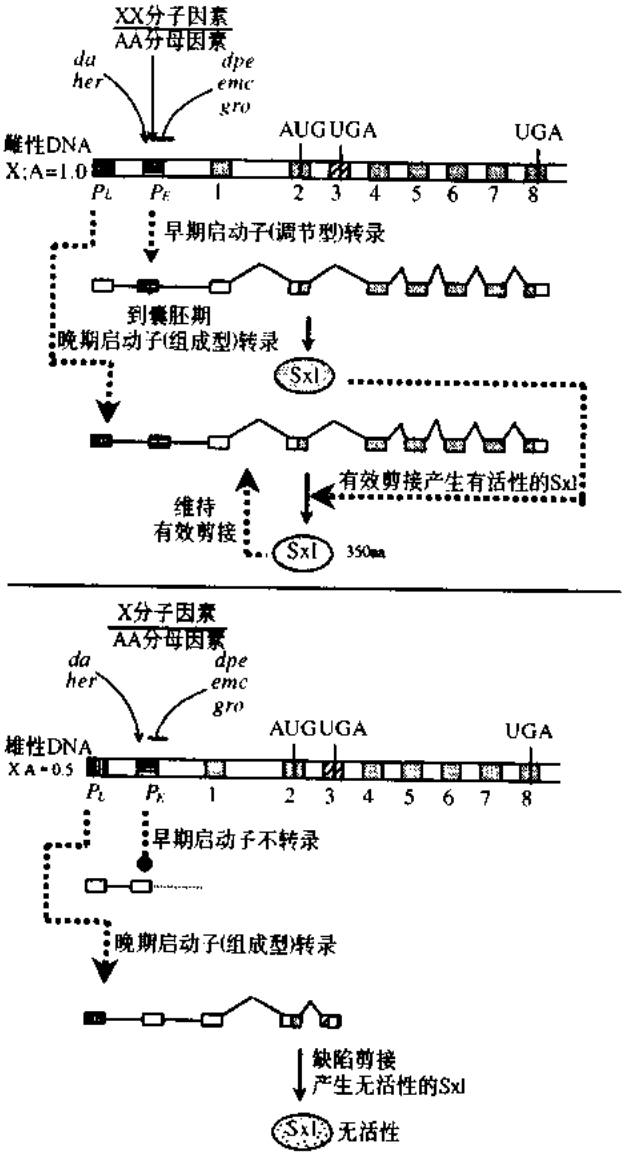


图 24 - 39　主开关基因 *Sxl* 在雌性果蝇(上)和雄性果蝇(下)中的表达与调节。

母因子的调控、激活，所以可在两种性别的囊胚形成后开始表达，Sxl 蛋白含有一定浓度的碱性氨基酸，是一种 RNA－结合蛋白，它可以改变晚期启动子产生的 Sxl 初始转录本的剪接，在雌果蝇中由于已存在早期启动子合成的 Sxl，因此可以对晚期启动子的 mRNA 进行适当剪接，从而能产生有活性的 Sxl，这就是 Sxl 的自我调节的功能。这样Sxl 的活性在雌体中得到了维持。

雄体（ X ： A = 0.5 ）其 Sxl 早期启动子无法被激活，也就不能产生 Sxl 蛋白，而晚期启动虽能启动，但初始转录物不能得到 Sxl 的适当剪接，第 3 个外显子也被转录，这样在起始点不远的位点就出现了终止密码，这便是雄性特异性的 mRNA ，只能产生一个很短没有功能的 Sxl 多肽。

在雌体中 *Sxl* 的自我反馈调节维持 *Sxl* 的正常表达起到了重要的作用。但还有其他的基因也涉及到维持 *Sxl* 的表达，如男性化基因 *vir*（*virilizer*），老处女基因 *shf*（*sahsfille*），它们突变后 *Sxl* 因不能维持高水平的表达而阻滞雌性发育。还有一个母性效应基因叫做雌性致死 *fl*(2) *d* （ female-lethal-2d，）基因。其突变会影响到 *Sxl* 晚期转录的雌性特异剪接。

4. *Sxl* 控制体细胞性别分化基因

在果蝇性决定的途径中在 *Sxl* 下游的还有 *tra*，*tra-2*，双重性别基因 *dsx*（double sex ），间性基因 *ix* （ intersex ），两性基因 *her* （ hermaphrodite ）和不育基因 *fru* （ fruitless ）等。

Sxl 直接的靶基因是 *tra* 基因， *tra* 在两性的发育中均被转录，但只在雌果蝇中产生有功能的 Tra 蛋白， Sxl 蛋白的存在可以改变 *tra* 基因转录本的剪接。一种剪接方式在雄性中都存在，那就是没有 *Sxl* 的指导，产生的 RNA 的前部含有终止密码子（图 24－40 雄性途径上图）。最终产生的 Tra 蛋白是没有功能的。在部分雌体细胞中Sxl 蛋白的存在与 *tra* mRNA 前体相作用，结合在非性别专一性 3′端剪接位点附近抑制了通常使用

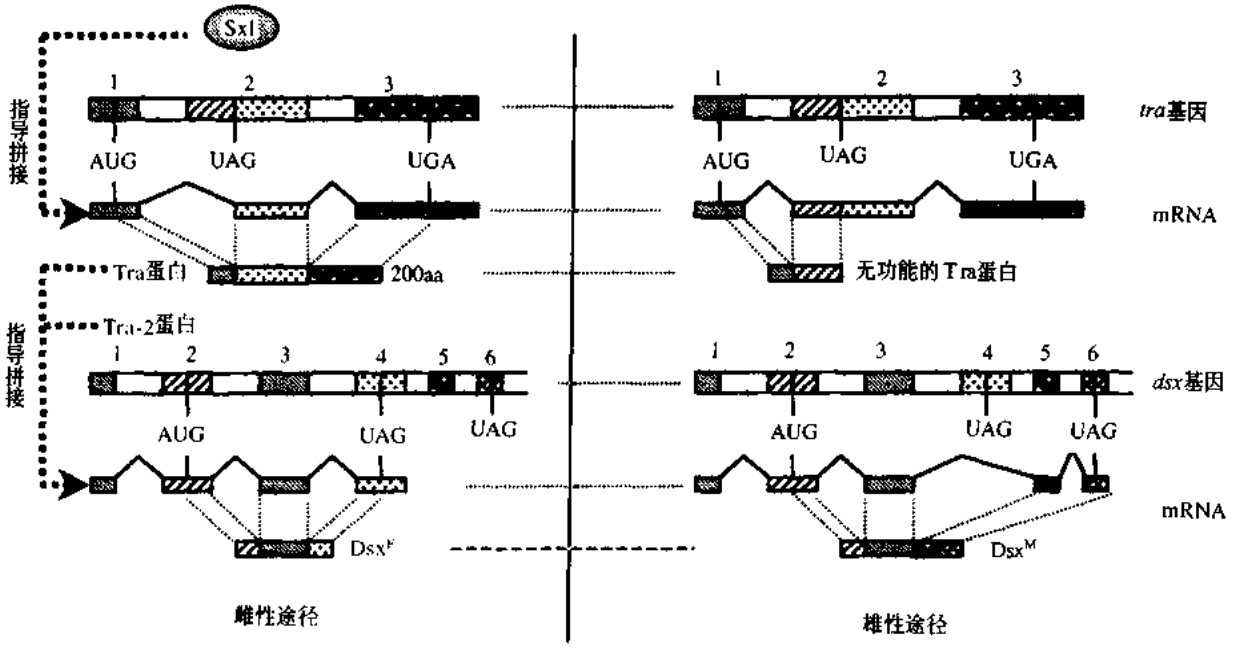


图 24 － 40 果蝇性别分化过程中 *Sxl* 、 *tra* 和 *dsx* 基因以 RNA 选择性剪接的方式进行级联调控。

的非性别专一的3′端剪切位点，当这个位点被跳越过时，便使用下一个第3个外显子中含有的UGA，结果使约一半 *tra* mRNA前体产生雌性特异性剪接，形成包括全长ORF的成熟的mRNA，这种mRNA编码有活性的Tra蛋白。

tra-2 基因有多个起始位点和多聚腺苷加尾位点，加上选择性剪接，就能产生很多不同的相关RNA，编码一系列相关蛋白，这些蛋白都含有Arg-Ser富集区和一个RNA识别区域（RRM），这是几个已知的作为RNA剪接因子的RNA结合蛋白的特点之一，表明tra-2的功能也与RNA剪接有关。有些 *tra-2* RNA只存在雄性生殖细胞中，为正常精子发生所需，和性别决定无关。另一些 *tra-2* 的产物在两性的体细胞中都存在，无性别差异，而只有在Tra蛋白存在对Tra－2蛋白才显示出影响性别分化的功能。Tra可能通过与Tra-2形成某种复合体的形式共同作用，通过对 *dsx* 的mRNA前体的雌性剪接，表达雌性特异的Dsx^F，从而进一步控雌性化发育。

dsx 座位有两种彼此排斥的表达方式，在雌性中在Tra和Tra-2的作用下，*dsx* 的mRNA初始转录本中从内含子的5′到其3′端被切除，结果翻译终止于外显子4末端的UAG处，形成一个较小的蛋白（图24－40）DSX^F；在雄性中Tra蛋白无功能，Tra-2不能与其形成复合物，*dsx* 的剪切也就不受二者的影响，从内含子3的5′端位点到内含子4的3′端被切除，这样将含有终止密码子的外显子4也切除了，使翻译一直进行到第6个外显子，产生一个较大的DSX^m蛋白。Tra和Tra-2调节的作用和 *Sxl* 有2点不同：①前者使剪接的结果存在中途的终止位点，而后者相反切除了中途终止位点，使蛋白有功能；②Tra和Tra-2在两性中不同的剪接虽也产生不同的蛋白，但都有活性，仅功能不同。而 *Sxl* 的作用结果是产生有活性的 *Sxl* 和Tra，而没有 *Sxl* 指导的剪接，*Sxl* 和 *Tra* 都没有活性。

两性基因 *ix*（interser，）目前了解不多，经遗传分析表明 *ix* 产物可与Dsx相互作用，在雌蝇中促使DSX^F对雄性分化起抑制作用，是雌性分化所需的基因。*her* 基因是母体温度敏感基因，它在果蝇性别决定途径的前、后两个位置上，都起到调节作用。其本身是一种具有锌指结构的转录因子。在途径的后部与 *ix* 作用的位置相同（图24－34）。

雌、雄果蝇在某些身体的结构，行为和生化等方面都具有明显差异。如雄果蝇的前腿上有**性梳**（sex comb），腹部末端有很宽的色素沉积带；在第五腹节有雄性特异的劳伦斯肌肉（musele of Laurence，*MOL*）；雄性果蝇还有特殊的求偶行为，这些特征都由不同的性别特异基因控制。不久前人们还认为性别特异基因是由 *dsx* 控制，但Taylor（1992）发现 *MOL* 的发育并不是由 *dsx* 控制的，而是由性别决定途径中居于上游的基因控制的。这表明在雄性中，紧接着 *tra* 和 *tra-2* 下游存在着一条分枝途径，此分支途径的前部基因是不育基因 *fru*（fruitless），*fru* 含有与在dsx中发现的相似的调节序列，*fru* 的突变显示了 *MOL* 缺掐。*MOL* 和雄性求偶行为都是受神经系统控制的，而有趣的是 *fru* 分支控制了神经系统功能这一性别特异方向，*fru* 突变体也出现了雄性求偶行为的缺陷。

5. 剂量补偿作用

在所有XY型的生物中，由于雌性有两条X染色体，而雄性仅一条X染色体，那么按理说X连锁基因表达的量雌、雄之间是不同的，雌性所具有的X－连锁基因的产量是雄性的2倍，但实际上是相近的，这就是通过一种剂量补偿机制来调节的，

但不同的生物这种剂量补偿机制也是不同的。

果蝇的剂量补偿效应仍通过 *Sxl* 蛋白来调节的（图 24 - 36）。有一组基因称为雄性特异致死效应基因（male-specific lethal effects），这组基因已发现有 4 个：*msl-1*、*msl-2*、*msl-3*、和 *mle*(maleless)。前 3 个基因都位于第 2 号染色体上，仅 mle 在第 3 号染色体上。用它们的抗体进行免疫反应，表明 Msls 蛋白在雄性的 X 染色体上有数百个结合位点，而在雌性 X 染色体上却没有。表明 *msls* 可能直接介导剂量补偿，调节同一条 X 染色体上性连锁基因。这 4 个基因的产物可能相互结合成异聚复合体。

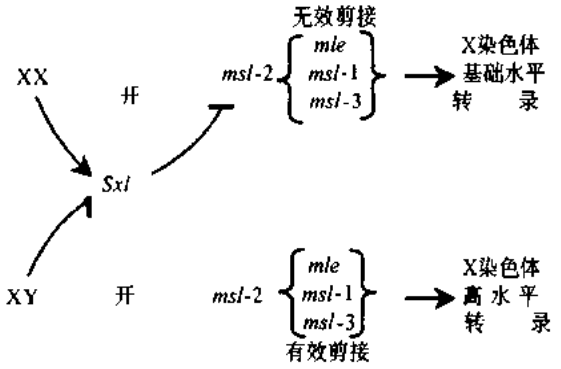


图 24 - 41 果蝇中 *Sxl* 基因对计量补偿的调控。

一系列证据表明 *Sxl* 基因通过阻止雌性 X 染色体的过度转录来调节剂量补偿。*Sxl* 功能的打开和关闭就像控制体细胞分化那样控制剂量补偿。近期用特异的多克隆抗体与 X 体色体的结合实验研究了 *Sxl* 的表达，结果表明 *Sxl* 在细胞中表达的水平与 *msl* 和 X 染色体结合能力相一致，这就意味着 *Sxl* 的表达可能促使 *msl* 与 X 染色体的结合，从而抑制了 X 染色体基因的过度表达。

msl-2 是作为剂量补偿的开关基因。也是 *Sxl* 调节目标的候选者。Msl-2 蛋白只在雄果蝇中存在，若在雌体中意外表达，则导致了另一些 Msls 广泛地积聚在 X 染色体上，彼此相间排列。

Sxl 蛋白也是通过选择性剪接的机制对 msl-2 进行负调节。在雄性中无 Sxl，*msl-2* 转录本 5′上游区域的一个内含子被切除，而在雌性中 Sxl 正好结合在此 5′上游内含子上，使其得以保留，而这个内含子上有终止密码子，使翻译过早终止，产生无功能的 Msl-2。

Msls 的调控是在转录水平，通过促进雄性 X 染色体转录水平来提高基因产物的量。

整个的果蝇 X - 连锁基因的剂量补偿调控如图 24 - 41 所示 X ： A 信号和母体基因在雌性中激活 *Sxl*，而 *Sxl* 使 *msl-2* 的 RNA 不能适当的剪接，结果产生了无效的 Msl-2 蛋白，从而就不能激活 msls 基因，X 染色体则以基础水平转录；而在雄性中 Sxl 蛋白无功能，msl-2 可不受干扰的正确剪接，产生有活性的 Msl-2，它可激活 *msls* 基因，使 X 染色体高水平转录，这样雄果蝇虽然只有一条 X 染色体，但表达的量与雌果蝇二条 X 染色体表达的量相同，达到了剂量平衡。

由于 *msls* 基因位于常染色体上，当常染色体套数增加 *msls* 的产量也随之增加，对 X 染色体表达的促进作用也更强，因此超雄（1X ： 3A）X 染色体的活性比正常的雄性（1X ： 2A）更高。

二、哺乳动物性别决定和分化

早在 1940 年 Ajost 等就通过一系列胎兔阉割实验搞清了哺乳动物性别决定和分化的基本概念，认识到性别的决定等同于睾丸的决定，即**初级性分化**。一旦性腺开始分化由其产生的激素如抗副中肾激素（antitimullerian hormone，AMH）和睾酮将诱

导体细胞表型的细胞分化，即次级性分化。直到20世纪60年代前后人们才弄清哺乳动物睾丸的决定是依赖于Y染色体，推测这是由于Y染色体上有一个基因，它编码了一种睾丸决定因子（人类：testis - determiningfactor, *TDF*；小鼠：terstis - drterminninggeneonthey, *Tdy*），决定性嵴发育成睾丸。人们一直在努力探寻 *TDF/Tdy* 基因的候选者，直到20世纪80年代末 PalmerM.S 等（1989）通过3例46XX男性和1例46XX真两性畸形的研究，发现他们的基因组中具有35kb的Y特异DNA序列的存在。1990年 Goodfelow 等进一步克隆了这一片段，并鉴别出一个开放阅读框，称 *SRY*（Sex-determiningregionY），在小鼠中称 *Sry*，以此区别。以下的证据有力地支持 *SRY/Sry* 就是 *TDF/tdy*：①经测序分析，发现 *SRY* 含79个氨基酸的基序，且和高泳动蛋白（highmobilitygroup, HMG）同源，具有DNA结合蛋白的特点，表明SRY可能具有调控的功能；②通过分子杂交发现在其他的哺乳动物中 *SRY* 仅在雄性中存在，且位于Y染色体上；③小鼠 *Sry* 表达的时空性符合 *Tdy* 对睾丸的决定作用，即 *Sry* 仅在受精后10.2 ~ 12天（*days* postcoitus, DPC）短期特异性表达，此期正值性嵴开始向雄性性腺分化之时，在雌性胎鼠性嵴细胞中无 *Sry* 的表达；④小鼠Sry基因定位于Y染色体上，小鼠 *Tdy* 突变体发育为XY雌性，用 *Sry* 为探针进行DNA杂交，发现这种XY雌鼠缺失了 *Sry* 序列。相反XX性反转小鼠可能是由于Y染色体片断易位到X染色体上所致，经分子杂交证实易位片段上确实具有 *Sry* 序列存在；⑤在46XY女性的 *SRY* 编码序列中曾发现过错义，无义和移码突变，除一种以外其余都发生在HMG的DNA结合区。例外的突变位点发生在HMG box的终止密码上，也有的是 *SRY* 侧翼基因序列发生突变所致；⑥最有力的证据是Koopman等将一段含有 *Sry* 的14kb DNA片段作为转基因导入XX雌性胎鼠中，结果部分转基因鼠产生了性反转、睾丸发育。故20世纪90年代初普遍认为 *SRY/Sry* 是睾丸形成的必要条件和充分条件。

人类的 *SRY* 基因含有一个单个的外显子，长850bp，在起始密码子的上游91bp处有一个主要的转录起始位点（图24-42b），在SRY-ORF 5′端4.0bp处有一个790bp的GpG岛，可能是Wilm's肿瘤基因产物WT1的潜在识别位点。意味着WT1可能直接或间接地影响到人类的性别决定。

小鼠的 *Sry* 也只有单个外显子，位于一个2.8kb的单一序列的延伸部分，侧翼是一个至少15kb的反向重复序列（图24-42a）。人类 *SRY* 和小鼠 *Sry* 的HMGbox功能区有71%的同源性，但此区外侧无明显同源性。*Sry* 在小鼠的初始性嵴中的转录本是

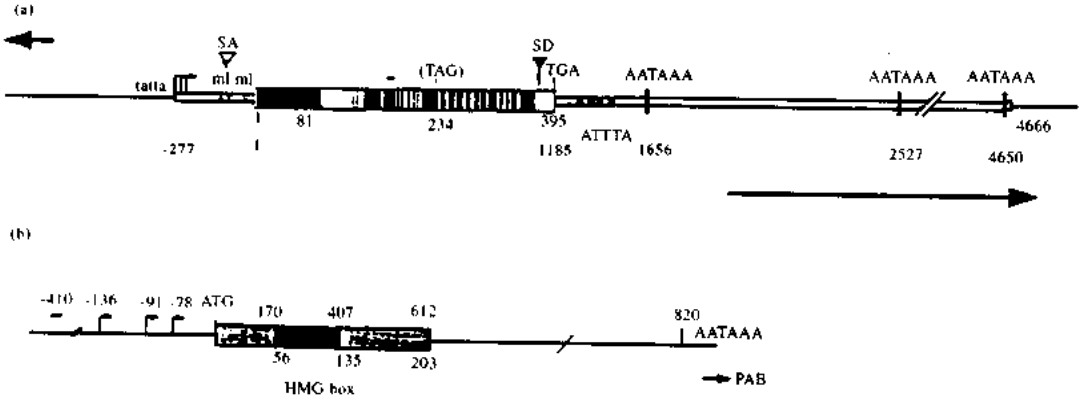


图24-42　小鼠 *Sry* 基因(a)和人类 *SRY* 基因(b)的结构。

呈线性的，但在成体睾丸中其转录本是以一种茎环状的 RNA 分子存在（可能由于侧翼反向重复序列所致），因此不和多核糖体相联而不能被翻译。表明 *Sry* 在成体睾丸中可能没有功能。人类 *SRY* 的转录本也可在成体睾丸中检测到，但未发现茎环状转录本。

SRY 和其他含有单个 HMG box 的蛋白一样以序列特异方式和 DNA 结合。人类的 SRY 显示出对 AACAA(A/T)(G/C)序列的识别、结合，并使 DNA 弯曲超过 70°，它可能作为转录的反式激活因子。实验表明 *SRY* 通过小沟接触 DNA。小鼠的 *Sry* 以高度的亲和力和 CATTGTT 结合，结合位置在大沟。

小鼠的 *Sry* 有一个 DNA 结合区和转录激活区。将 GAL4 DNA 结合功能区（1～98 密码子）和 *Sry* 基因的不同区域重组，构成嵌合基因，再将这种嵌合基因和含有 GAL4 结合位点的受体基因共转染 CHO 细胞，结果表明小鼠的 Sry 的 C 端具有转录激活区，而同样用人类的 *SRY* 来构造嵌合基因，在此系统中都不能激活转录，表明人类的 SRY 可能有一种抑制作用。

已知 *SRY*-ORF 内的突变可导致 46XY 性反转为女性，但近年来发现这种突变有时并未引起性反转。有 4 个家族有 46XY 性反转，患者 HMG DNA 结合区发生了突变，但患者们的父亲也有同样的突变却未出现性反转。这就意味着性别决定是复杂的，可能还有另一些基因也介入了性别决定和分化。在 46XX 性反转男性中虽可测出 *SRY* 的易位片段，但个体表型存在着一个变化范围，这也需作出解释。

近年来又发现了一些有悖于 *SRY*/ *Sry* 决定睾丸发育的报道：①人类家族性 XX 性反转男性的某些病例并不存在 *SRY* 基因，但有的可发育成正常的男性；②已报道 15 例带有两性或女性外生殖器的 XY 患者发现有 Xp 部分重复，但 *SRY* 基因都是明显正常的；③南美仓鼠 *Akodone* 有两种 Y 染色体，其中一种 Y 染色体传递给小鼠后，诱导睾丸发育长成雄鼠；而另一种 Y 染色体却不能诱导子鼠的性腺分化，从而产生了不育的 XY 雌鼠。对这种性反转鼠进行了 *Sry* 测定，发现并非是 *Sry* 的缺失或突变所致，而是在性分化阶段 *Sry* 未能表达；④在 *Ellobius* 属的鼹－田鼠（mole-vole）两个种中是没有 *Sry* 的，但性别依然可以分化；⑤ *Talpa* 属所有雌鼠都没有 Sry，性腺却发育成卵睾（ovotestes）。以上表明 *SRY*/*Sry* 仅是涉及性别决定过程的基因之一，还有一系列基因（*WH1*, *SF-1*, *SOX9*, *AMH*, *OAX*, *DSS* 等）可能也参与了性别决定和性腺的发育。

最近已发现 *WT1*， *SF1* 和 *SOX9* 三个基因参与了哺乳动物性别决定。这三个基因都在未分化的性腺中进行表达；而且它们的产物有序列特异 DNA －结合功能域和蛋白质结合区。

Wilm's 肿瘤抑制基因 *WT1* 的 DNA 结合区有 4 个 Cys_2His_2 锌指结构，与转录因子 EGR 家族有很强的同源性。此基因位于人类 11 号染色体上（小鼠为 2 号染色体上），当 *WT1* 发生突变导致肾小球肾病和性腺发育异常，某些位点突变的杂合体显示出 XY －性反转的表型特征。这表明 WT1 可能在性别分化的早期参与了调节作用。将 *WT1* 接上报告基因，然后共转染 N1H－3T3 细胞，结果表明此蛋白的 DNA－结合区的外侧具有很强的转录阻遏作用，因此推测 *WT1* 的作用可能是一个转录阻遏物。

SF-1（steroidogenic factor-1）已知是一种核内受体蛋白基因，位于小鼠的第 2 号染色体上，*SF-1* 具有间隔开的配基结合区和 DNA 结合区，它在固醇羟化酶的调节中起着重要的作用。*SF-1* 在性腺发育和性别决定中具有一定的作用，这有三方面的证据：①在受精后 9～9.5 天鼠胚的尿殖嵴中可测到 *SF-1* 的转录本，也可在胚胎的支持细胞（或称足细胞，Sertoli's cell）中测到其表达；②*SF-1* 显示出对 *AMH*（抗副中肾激基基因，也

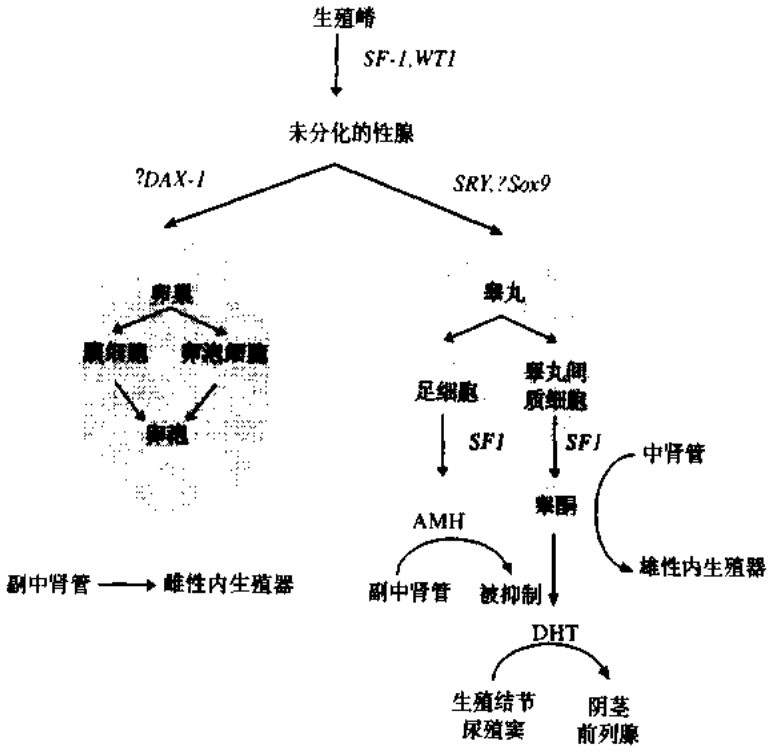


图 24 - 43　哺乳动物性别决定中的相关基因的功能。

称副中肾抑制基因，mullererian inhibiting substance，*MTS*）具有直接的调节作用。在 *AMH* 基因的启动子上有 *SF-1* 的结合位点，且它对 *AMH* 基因的激活是必须的；③用剔除（knockout）技术去除小鼠体内的 *SF-1* 基因时小鼠的肾上腺和性腺缺如，*SRY*，*SF-1* 和 *AMH* 三者之间的关系如图 24 - 43 所示。

Sox9（ *SRY* - 相关基因， *SRY*-relatedbox ）是与性腺发育有关的第二个早期基因，位于人类的第 17 号（小鼠的 11 号）染色体上，近年已被克隆，它和 SRYHMG box 区域的同源性大于 50%，突变时可导致产生侏儒综合征（如 Campomelic dysplasis，CD 症患者出现很多的骨骼异常，且 30 %以上带有不同程度生殖器缺陷和 46XY 性反转。*Sox9* 基因编码一个 DNA 结合蛋白的 HMG 超家族成员，它可能和 *SRY* 在转录调节中有着相似的功能，具体机制尚不清楚。

MIS（ 即 *AMH* ）定位在人类的第 19 号染色体上（小鼠的 10 号染色体）。其产物 MIS 在形态分化前正在发育的睾丸中首先产生的。它是在支持细胞中合成的。MIS 是 TGF-β（转生长因子，transforming growth factor）超家族中一种同形二聚体糖蛋白， 140kDa 。二聚体的翻译后水解激活了激素，但两个多肽片段保持互相连结，甚至 C 末端 25kDa 的二聚体片段也能独立地使副中肾管退化。*MIS* 发生突变使得 46 XY 男性体内存在子宫和输卵管。

DAX-1（ *DSS-AHC* critical region on the X，gene 1 ）在 X 染色体上，在性别决定途径中发挥重要作用的 *dax-1* 基因中已鉴别出 *SF-1* 的结合位点。*DAX-1* 突变可导致先天性肾发育不全（ adrenal hypoplasia congenita，AHC ），这种性连锁疾病常伴有性成熟缺陷。表明 *DAX-1* 和性腺结构的发育有间接的关系。当 *SF-1* 发生无效突变时肾上腺或性腺都不能发育，而 *DAX-1* 突变仅导致肾上腺在胎儿阶段的发育受阻，这表明 *DAX-1* 的影响可能在发育途径中处于 *SF-1* 作用的下游。

DSS（剂量敏感性反转基因，dosage sensitive sex reversal）基因是 Xp21 的重复保守序列，这个基因与 *AHC*（adrenal hypoplasia）、*HHG*（hypogonadotrophic hypogonadism）连锁于 X 染色体上。在正常情况下，此基因为 *SRY* 所抑制，但当它有两个重复的活性拷贝时，可导致 *SRY* 正常的 *XY* 个体发育成女性。通过进一步研究，从 *DSS* 区域中克隆了 *DAX－1* 基因，它编码核内激素受体家族的一个新成员。在鼠胚 11.5dpc 时开始表达，这和性别决定的时相是一致的。

近期新提出的一个人类性别决定和分化的模型（图 24－44），认为单拷贝的活性 DSS 能在正常的 XX 中可阻遏雄性途径的发育，但在 XY 雄性中则受到 *SRY* 基因的作用而失活。在 XY 个体中，雄性途径基因 *MPGs* 可在正常睾丸分化时被激活，这是由于 DSS 单拷贝基因受到 *SRY* 的抑制。在正常睾丸分化时时期的激活将导致正常雄性表型的发育。在 XX 雌体中无 *SRY*，这样使 *DSS* 得以维持 *MPGs* 的抑制。雌性途径基因（*FPGs*）在发育中较晚一些被激活。各种情况表明转分化仅在卵母细胞衰竭后才发生。卵巢分化后 *MPGs* 不再受到 *DSS* 的抑制，而是受到卵母细胞的抑制。

以上模型较合理地解释了目前发现的各种性反转现象，但哺乳动物性别决定的研究尚未达到终点，许多问题仍有待进一步的研究。

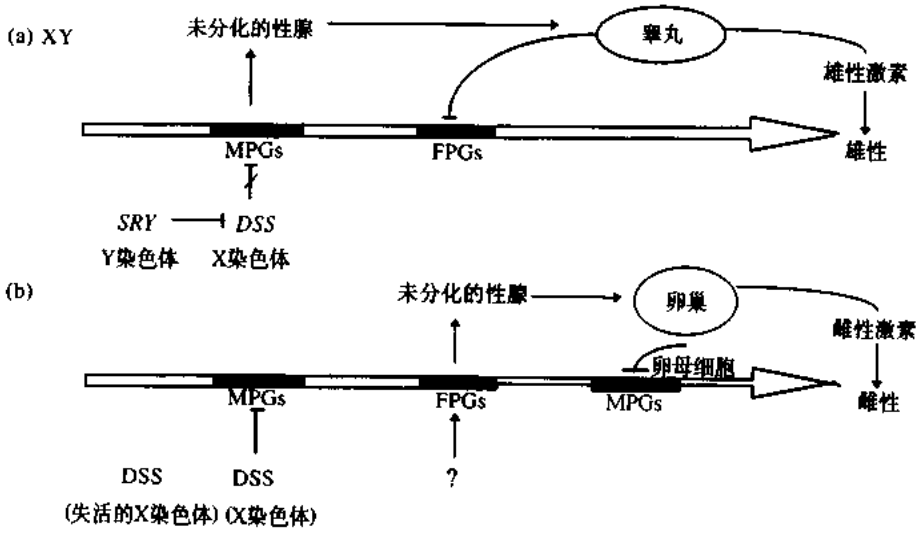


图 24－44 哺乳动物性别决定和不同类型 XX 和 XY 性反转模型 MPGs：雄性途径基因；FPGs：雌性途径基因。
（转摘自：Jimenes R，Sanches A，Burgos M *et al*.：TIG 12：164(1996)）

第二十五章

群体遗传学

基本原理

群体的基因库是一群有性生殖的个体中所有基因的总和。基因库能用术语基因频率和基因型频率来描述。

哈迪－温伯格定律是在一个无限大而随机交配的群体中，没有进化因素存在时基因频率不会改变。当群体的基因频率是 p 和 q，基因型频率为 p^2，2pq 和 q^2 时，经过一代以后基因型频率达到稳定。

经典模型、平衡模型和中性突变模型都是用来解释在自然群体中有多少遗传变异的一些假说。并解释产生遗传变异的原因。

突变、迁移和自然选择是可以改变群体中基因频率。

突变可改变基因频率，在没有其他因素存在时，正向突变和回复突变的相对速率能否达到平衡取决于群体基因频率。

遗传漂变是由于样本机误随机改变基因频率。遗传漂变使群体中的遗传差异发生改变，丢失某些遗传变异。

遗传漂变是一种分散性效应，群体基因频率变动的大小可以确定，但变化的方向不确定。样本越小漂变越厉害。

迁移也叫基因流，涉及到群体基因的移动。迁移能改变群体的基因频率，倾向于减少遗传趋异。

自然选择是基因型的差异繁殖。它可通过适应值来度量。适应值是基因型的相对繁殖的能力。自然选择对群体的基因库可产生很多不同的效应。

非随机交配影响到群体的基因型频率。近交使纯合体增加，远交使杂合体增加。

进化速率能通过 DNA 或 RNA 的顺序的比较来量度。不同的基因甚至相同基因的不同部分以不同的速率趋向演变。

在真核生物中有些基因以多拷贝存在，这些基因的顺序相同或相似。一组这样的基因称为多基因家族。

一些生物 mtDNA 进化速率比核 DNA 要快一些。

一致进化是指在一个物种中相同顺序的多拷贝之间要维持顺序的一致性。

生物中的进化关系可以通过研究 DNA 和 RNA 的顺序而得以揭示。

遗传学可以进一步分成**传递遗传学**（transmission genetics）（也可称为经典遗传学）、**分子遗传学**和**群体遗传学**（population genetics）三个主要的亚学科。这三种亚学科是在不同的历史背景下形成的。传递遗传学是在孟德尔定律发现以后建立起来的，它主要的内容是研究基因纵向传递的规律。它研究的对象是个体。分子遗传学开始于大分子结构的研究。人们对 DNA 怎样编码遗传信息，以及在细胞中这些遗传信息又怎样变成为表型进行系统地研究和探讨。分子遗传的一系列过程集中在细胞中。

群体遗传学是研究一群个体的遗传规律。它涉及到群体中遗传变异的方式，这些方式怎样发生改变或进化的。群体遗传学研究的对象不再是细胞或个体，而是**孟德尔式群体**（Mendelian population）。孟德尔群体是由一群可杂交繁殖的个体组成，这些个体具有共同**基因库**（gene pool）中某些基因。这种基因库也就由孟德尔群体中的各个体的基因组成。为了了解进化过程中的遗传学，我们就必须要研究孟德尔式群体的基因库，而不是单个成员的基因型。各个个体除了突变以外，基因型是终身不变的，那么进化又

是怎样产生的呢?

群体遗传学需要研究是基因在群体中的遗传行为。它涉及到对稳定和变化的环境的研究,也涉及到进化机制的研究。进化论虽可以从很多不同的角度,用各种不同的方法加以研究,但归根结蒂,它就是一个基因频率变动的过程。所以群体遗传学的中心任务就是研究基因频率变化的动力学。

群体遗传学的理论基础是建立在遗传学和进化论上。在20世纪30年代由三位杰出的学者在以上基础上建立群体遗传学,他们是R.A.Fisher(1928年出版了《自然选择的遗传学理论》)、S.Wright(1931年在Genetics刊物发表了著名论文《在孟德尔式群体中的进化》)及J.B.S.Haldame (1932年出版了《进化的原因》一书)。

第一节 基因型频率和基因频率

一、基因型频率

为了研究孟德尔式群体的遗传组成,群体遗传学必须首先讨论群体的基因库。这就要计算群体中的**基因型频率**和**等位基因频率**。以百分比来表示的频率总是在0~1之间的区域内。比如高加索人中有43%的人是红色头发,那么在这个群体中红发的频率就是43%。为了计算一个特异基因座的基因型频率,我们可以统计群体中带有一种特殊基因型的个体数。这个特的座位上各种基因型频率之和应为1。以深红虎蛾(*Panaxia dominula*)(图25-1)翅上斑点的样式为例,在大部分群体中存在3种基因型,而每种基因型的表型都不同。E.B.Ford在英国收集了497只虎蛾,统计它们的基因型 分别为:*BB* = 452、*Bb* = 43和 *bb* = 2。它们的基因型频率分别是

P = f(BB) = 452/497 = 0.909
H = f(Bb) = 43/497 = 0.087
Q = f(bb) = 2/497 = 0.004
合计:1.000

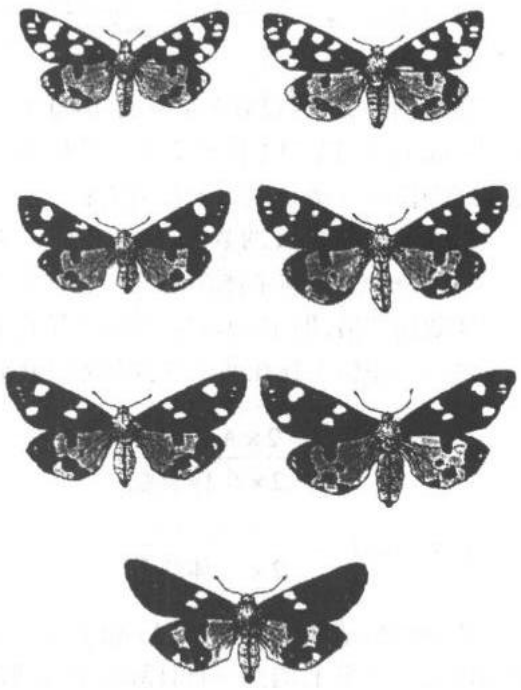


图25-1 深红虎蛾(*Panaxia dominula*)翅上斑点的样式。顶部的两种蛾子是正常的纯合体(*BB*)。第2,3排是杂合体(*Bb*)。底部的蛾子是异常的纯合体(*bb*)。(引自Russell,.P.J.:Genetics, 3rd ed., 1992, Fig.23-1)

群体遗传学常用P,H和Q三个字母来表示带有2个等位基因座位上的3种不同的基因型。小写字母 p 和 q 是用来表示等位基因的频率。应注意,符号不能乱用。在人类群体中对于不同血红蛋白变异体的基因型频率也用相同的方法计算。

二、基因频率

基因型频率常在某些进化压力作用于一个群体时被使用，而在大部分情况下群体遗传是用等位基因来描叙基因库。在某一个座位上等位基因的频率被称为**等位基因频率**(allelic frequency)，或称为**基因频率**(gene frequency)，后者使用更为普遍。基因频率的使用常优于基因型频率。首先等位基因总是比基因型少一些，这样使用基因频率基因库就可以用较少的参数来描述。若一个座位有 3 个等位基因，那么就要用 6 种基因型频率来描述基因库，若用基因频率的话用 3 种就可以了。再者在有性生殖的生物中形成配子时，配子只含等位基因，而并无基因型。在代代相传的过程中只有等位基因是连续的，而基因库的进化也是通过等位基因频率的改变。

基因频率可以用两种方法来计算：一种是通过获得的不同基因型的数目，另一种是通过基因型频率。首先我们可以从基因型数来直接计算基因频率。

$$\text{基因频率} = \frac{\text{群体中某个座位的特定等位基因的拷贝数}}{\text{群体中该座位所有等位基因数}}$$

现举例说明。假设有个群体有 1000 个二倍体个体，其中 353 个为 AA，494 为 Aa 和 153 为 aa，每个 AA 型个体有 2 个 A 等位基因，而 Aa 杂合子中只有一个 A 等位基因，因此在此群体中 A 等位基因的数应为：

2 × AA 的数 + Aa 的数 = (2 × 353) + 494 = 1200。

由于每个 2 倍体个体在此座位上有两个等位基因，所以群体中等位基因的总数应是个体数的 2 倍，即 1000 × 2。用这个等式来表示基因频率是 1200 ÷ 2000 = 0.6

当一个座位上存在两个等位基因时我们可用下列方式来计算基因频率：

$$p = f(\mathrm{A}) = \frac{2 \times \mathrm{AA} + \mathrm{Aa}}{2 \times \text{个体总数}}$$

$$q = f(\mathrm{a}) = \frac{2 \times \mathrm{aa} + \mathrm{Aa}}{2 \times \text{个体总数}}$$

两个等位基因的频率，$f(\mathrm{A})$ 和 $f(\mathrm{a})$ 一般以 p 和 q 来表示。一个座位上的基因频率相加总是等于 1，因此一旦计算出了 q，那么 $p = 1 - q$。

三、复等位基因的基因频率

假设现在在一个座位上有 3 个等位基因 A^1，A^2 和 A^3，我们想确定这些基因的频率。我们可采用与二个等位基因相同的方法，即每种等位数除以群体中这个座位上等位基因的总数：

$$p = f(\mathrm{A}^1) = \frac{(2 \times \mathrm{A}^1\mathrm{A}^1) + (\mathrm{A}^1\mathrm{A}^2) + (\mathrm{A}^1\mathrm{A}^3)}{2 \times \text{个体总数}}$$

$$q = f(\mathrm{A}^2) = \frac{(2 \times \mathrm{A}^2\mathrm{A}^2) + (\mathrm{A}^1\mathrm{A}^2) + (\mathrm{A}^2\mathrm{A}^3)}{2 \times \text{个体总数}}$$

$$r = f(A^3) = \frac{(2\times A^3A^3) + (A^1A^3) + (A^2A^3)}{2\times 个体总数}$$

为了说明一个座位上存在2个以上等位基因时基因频率的计算，现用牛奶草甲虫遗传变异的研究资料为例：Walter Eanes 和他的同事们测定了编码葡萄糖磷酸变位酶(PGM)这一座位上的基因频率。此座位上发现有3个等位基因，每个等位基因编码了此酶的不同分子变异体，在一个群体中基因型的数目收集如下：AA = 4, AB = 41, BB = 84, AC = 25, BC = 88, CC = 32, 共计274个甲虫。它们的等位基因频率是：

$$f(A) = p = \frac{(2\times 4) + 41 + 25}{(2\times 274)} = 0.135$$

$$f(B) = q = \frac{(2\times 84) + 41 + 88}{(2\times 274)} = 0.542$$

$$f(C) = r = \frac{(2\times 32) + 88 + 25}{(2\times 274)} = 0.323$$

在以上计算中某等位基因纯合体乘2，杂合体不乘2，相加以后除以2倍群体中具有此座位个体数。注意分子的部分不能将各种杂合体数都加上，而只能加带有我们所计算的那个等位基因的杂合体。用以上同样的方法我们可以计算出具有更多等位基因的基因频率。

四、由基因型频率来计算基因频率

第二种计算基因频率的方法是从基因型频率来计算基因频率。若我们已经确定了基因型频率，那用这种方法很快就可以计算出基因频率。如一个座位有两个等位基因，基因频率是

$$p = f(A) = (AA 的频率 + \frac{1}{2} Aa 的频率)$$

$$q = f(a) = (aa 的频率 + \frac{1}{2} Aa 的频率)$$

用纯合子的频率加上一半杂合子的频率是因为杂合子中一半是A，另一半是a。如果群体中有3个等位基因的话(A1, A2, A3)，那么它们的基因频率是

$$p = f(A^1) = f(A^1A^1) + \frac{1}{2}f(A^1A^2) + \frac{1}{2}f(A^1A^3)$$

$$q = f(A^2) = f(A^2A^2) + \frac{1}{2}f(A^1A^2) + \frac{1}{2}f(A^2A^3)$$

$$r = f(A^3) = f(A^3A^3) + \frac{1}{2}f(A^1A^3) + \frac{1}{2}f(A^2A^3)$$

虽然这种方法的计算比直接从基因型数目来计算要快，但容易产生误差，因此前面的方法更适用。

五、X－连锁座位上的基因频率

计算 X－连锁座位上的基因频率是较为复杂的，因为雄性只有单个的 X－连锁等位基因。但我们仍可以用计算常染色上基因频率相同的方法来计算。在雌性纯合体中带有两个相同的 X－连锁等位基因，而在雌性异合体中也只有一个特定的等位基因。而在所有雄性中只有一个 X－连锁等位基因。为了在 X－连锁座位上确定等位基因的数目，我们就得将雌性纯合体的数目乘 2 加上雌性杂合体数目和雄性半合体数目然后除以群体中等位基因的总数。确定等位基因总数时要用 2 倍的雌体数(因每个雌体都有两个 X－连锁基因座位)加上一倍的雄体数(因雄体只有单个 X－连锁基因座位)。

$$p = f(X^A) = \frac{(2X^AX^A♀) + (X^AX^a♀) + (X^AY♂)}{2\times 雌体数 + 雄体数}$$

例如有两个 X－连锁的等位基因(X^A 和 X)可用下列等式来计算：

$$q = f(X^a) = \frac{(2X^aX^a♀) + (X^AX^a♀) + (X^aY♂)}{2\times 雌体数 + 雄体数}$$

X－连锁座位上的基因频率也可由基因型频率来确定

$$p = f(X^A) = f(X^AX^A) + \frac{1}{2}f(X^AX^a)f(X^AY)$$

$$q = f(X^a) = f(X^aX^a) + \frac{1}{2}f(X^AX^a) + f(X^aY)$$

以上的等式关键是掌握其原理，原理清楚了也就不必去死记硬背它。

第二节　哈迪－温伯格法则

哈迪－温伯格法则(Hardy－Weinberg)是群体遗传中最重要的原理，它解释了繁殖如何影响群体的基因和基因型频率。这个法则是用 Hardy G H （英国数学家）和 Weinberg W(德国医生)两位学者的姓来命名的，他们于同一年(1908 年)各自发现了这一法则。他们提出在一个不发生突变、迁移和选择的无限大的相互交配的群体中，基因频率和基因型频率将逐代保持不变。

哈迪－温伯格定律可分为 3 个部分：第一部分是假设：在一个无穷大的随机交配的群体中，没有进化的压力(突变、迁移和自然选择)；第二部分是基因频率逐代不变；第三部分：随机交配一代以后基因型频率将保持平衡：p^2 表示 AA 的基因型的频率，$2pq$ 表示 Aa 基因型的频率 q^2 表示 aa 基因型的频率。其中 p 是 A 基因的频率；q 是 a 基因的频率。基因型频率之和应等于 1，即 $p^2 + 2pq + q^2 = 1$。这个定律简而言之：在没有进化影响下当基因一代一代传递时，群体的基因频率和基因型频率将保持不变。

一、理想群体

哈迪－温伯格定律的第一部分是前题，假设这些条件存在时此定律才适用。实际

上这些条件是不可能存在的，所以具备这些条件的群体称之为"**理想群体**"。首先，定律指出这个群体是无穷大的，若一个群体的大小有限，可能导致基因频率和预期的比例随机发生偏差。这种基因频率的改变就称**遗传漂变**(genetic drift)。在定律第一部分所谓的无穷大完全是设想的。没有任何群体具有无穷的个体。然而样本的误差仅对一个相当小群体的基因频率有明显的影响。实际应用时群体不需无穷大，只要不至太小即可。

哈迪－温伯格定律的第二个条件是交配必须是随机的。**随机交配**(random mating)是指各基因型之间的交配和群体中这些基因型的频率成正比。更为特别的是两个基因型之间交配的概率等于两个基因型频率的乘积。

为了说明随机交配，现以人类的 M－N 血型为例来解释。M－血型是由于在细胞的表面上存在一种抗原，与 ABO 系统的抗原相似。但在 M－N 系统中除了产生不相容以外，在输血时并不会产生凝血。M－N 血型是由带有两共显性等位基因 L^M 和 L^N 的座位决定的。在爱斯基摩人的群体中，3 种 M－N 基因型频率分别为 $L^M/L^M \approx 0.835$，$L^M/L^N = 0.156$，$L^N/L^N = 0.009$。若爱斯基摩人的婚配是随机的，那么 L^M/L^M 男人和 L^M/L^M 女人婚配的率就应等于 L^M/L^M 基因型频率乘以 L^M/L^M 的频率，即 $0.835 \times 0.835 = 0.679$。其他基因型之间的婚配率的计算也与此相似。

哈迪－温伯格定律所要求的随机交配常被曲解，误以为随机交配是针对所有性状。若是这样那么人类群体就不能符合哈迪－温伯格定律的要求了。因人类择偶并不是随机的，而是对智商、外貌、性格、身高、肤色、学历以及社会地位等都有一定的要求。虽然对某些性状的要求不是随机的，但大部分人对血型等并无要求，甚至有的人并不知道自己的 M－N 系统的具体血型。因此哈迪－温伯格定律要求的随机性是指诸如像血型这样一些性状，而不是那种非随机性状的座位。

第 3 个条件是没有进化的力量。在哈迪－温伯格定律中人们只关注遗传是否会改变基因频率以及繁殖怎样会影响到基因型频率。因此进化力量可被排除。后面我们仍要讨论进化及其对群体基因库的影响。在没有进化力量作用在群体上时，只适用于某些座位，其他的座位可能照样受到进化力量的影响。

二、哈迪－温伯格定律的预言

若满足了理想群体的条件，那么这个群体在遗传中是平衡的而且预期有 2 个结果。一是等位基因频率逐代不变，因此在这个座位基因库不会进化。二基因型频率将以 p^2，$2pq$ 和 q^2 的比例存在于随机交配的以后的各代中。群体的基因型频率以这个比率存在时就称为**哈迪－温伯格平衡**(Hardy－Weinberg equiliberum)。哈迪－温伯格定律的一个重要的应用是对达到平衡的群体来说从他们基因频率可以确定其基因型频率。

总之在理想群体中，哈迪－温伯格定律可以预计基因频率和基因型频率。哈迪－温伯格定律指出，基因频率将不会改变，仅在第一代中基因型频率发生改变。基因型频率将取决于基因频率。

三、哈迪－温伯格定律的由来

哈迪－温伯格定律说明当一个群体达到平衡的话，其基因型频率为 p^2、$2pq$ 和 q^2。为了了解在平衡时这些频率的基础假设等位基因 A 的频率是 p，等位基因 a 的频率

是q，在产生配子时，每种基因型都具有2种等位基因，它们的频率相等，因此在配子中A和a的频率分别是p和q。表25－1表明随机交配时的配子的组合。这个表说明等位基因频率和基因型频率的关系，这种关系形成了哈迪－温伯格定律的基础。我们可以看到配子随机配对时将产生的基因型频率为p^2(AA)、$2pq$(Aa)和q^2(aaZ)。这个结果来自等位基因频率等式的展开：$(p+q)^2 = p^2+2pq+q^2$，在随机交配一代以后基因型可以达到此比例。例如有一个群体其三种基因型频率为$f(AA)=0.2$，$f(Aa)=0.2$，$f(aa)=0.6$，那么两种配子中的基因频率为：

$$f(A) = f(AA)+\frac{1}{2}f(Aa) = 0.2+\frac{1}{2}(0.2) = 0.3$$

$$f(a) = f(aa)+\frac{1}{2}f(Aa) = 0.6+\frac{1}{2}(0.2) = 0.7$$

若随机交配，可求出下一代中基因频率为：

$f(AA)=0.09$；$f(Aa)=0.42$；$f(aa)=0.49$

我们再来计算下下一代配子中的基因频率

$$f(A) = 0.09+\frac{1}{2}(0.42) = 0.3$$

$$f(a) = 0.49+\frac{1}{2}(0.42) = 0.7$$

表25－1　随机交配的组合

		♂	
		p(A)	q(a)
♀	p(A)	p^2(AA)	pq(Aa)
	q(a)	pq(Aa)	q^2(aa)
总计 $p^2(AA)+2pq(Aa)+q^2(aa)=1$			

表25－2　随机交配，求出基因型频率

	A	a
	0.3	0.7
A 0.3	0.09AA	0.21Aa
a 0.7	0.21Aa	0.49aa

结果和上一代的基因频率相同，也就是说达到了平衡。

哈迪－温伯格定律还说明在理想群体中基因和基因型频率各代保持不变，表25－3中表示各种交配的可能性。随机交配的意思是两个基因型交配的概率等于两个基

表25－3　在随机交配的群体中遗传平衡的代数检测

交配类型 ♀ ♂	交配概率	后代频率		
		AA	Aa	aa
$p^2(AA)\times p^2(AA)$	p^4	p^4	–	–
$p^2(AA)\times 2pq(Aa)$ $2pq(Aa)\times p^2(AA)$	$4p^3q$	$2p^3q$	$2p^3q$	–
$p^2(AA)\times q^2(aa)$ $q^2(aa)\times p^2(AA)$	$2p^2q^2$	–	$2p^2q^2$	–
$2pq(Aa)\times 2q(Aa)$	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
$2pq(Aa)\times q^2(aa)$ $q^2(aa)\times 2pq(Aa)$	$4pq^3$	–	$2pq^3$	$2pq^3$
$q^2(aa)\times q^2(aa)$	q^4	–	–	q^4
总计	$(p^2+2pq+q^2)^2=1$	$p^2(p^2+2pq+q^2)=p^2$	$2pq(p^2+2pq+q^2)=2pq$	$q^2(p^2+2pq+q^2)=q^2$

基因型频率 $=(p+q)^2=p^2+2pq+q^2=1$；

基因频率 $=p(A)+q(a)=1$

因型频率的相乘积。如 AA × AA 的概率将等于 $p^2 \times p^2 = p^4$。其后代的频率也列于表中。我们可以看到 AA × Aa(或 $2p^3q$) 和 Aa × AA(或 $2p^3q$) 的交配概率为 $4p^3q$。而从孟德尔法则中可以知道这种杂交将产生$\frac{1}{2}$AA 和$\frac{1}{2}$Aa 的后代，因此得到 AA 后代的概率是：$4p^3q \times \frac{1}{2} = 2p^3q$。通过每种类型的交配所产生的后代的概率表中已列出。在表的最后一栏是每一种基因型的总计，此是通过相加而获得的。正如我们所能看到的，随机交配之后基因型交配频率仍为 p^2, $2pq$ 和 q^2, 而基因频率保持为 p 和 q。这个群体在合子和配子阶段存在的频率是：

合子：$(p^2AA) + 2pq(Aa) + q^2$ (aa)，配子 $p(A) + q(a)$

每一代合子以 p 和 q 的比率产生 A 和 a 配子。配子随机结合形成 AA，Aa 和 aa 合子，在群体中以 $p^2 + 2pq + q^2$ 的比率存在，这样周而复始，不断循环。

哈迪－温伯格定律指出，在平衡时基因型频率取决于等位基因的频率。二者之间的关系以图 25－2 所示。①杂合子的极大值为 0.5，而且只有当 A 和 a 的频率两者皆为 0.5 时 Aa 才有极大值；②若基因频率在 0.33 和 0.66 之间，杂合体的基因型频率最大；③当一个等位基因的频率降低时，其另一个等位基因的纯合体频率升高。

当 $f(a) = q = 0.1$ 时，$f(AA) = 0.81$, $f(Aa) = 0.18$, 而 $f(aa) = 0.01$(图 25－2)。

在人类中遗传病的分布也可说明这一点。对于一种罕见的隐性性状的基因频率要比性状本身的频率要高得多，这是因为大部分稀有基因是以不起作用的杂合状态存在。例如白化是人类中罕见的隐性性状。白化的一种类型是酪氨酸酶缺乏的白化。这类患者酪氨酸酶没有活性，而这种酶是合成黑色素所必需的。在北美白人中这种白化发生的频率为 1/40 000 或 0.000025。由于白化是隐性性状，患者的基因型为 aa。根据哈迪－温伯格定律 aa 基因型频率为 p^2。若 $p^2 = 0.000025$, 那么 $q = 0.005$, 而 $p = 1 - q = 0.995$, 所以杂合子的频率是 $2pq = 2 \times 0.995 \times 0.005 = 0.00995$ (近似于 1%)。这样当白化的频率很低时(1/40 000)，而其杂合体的频率却很高(近似 1/100)。可见即使某些性状是罕见的，但其杂合体的频率可能并不太低。

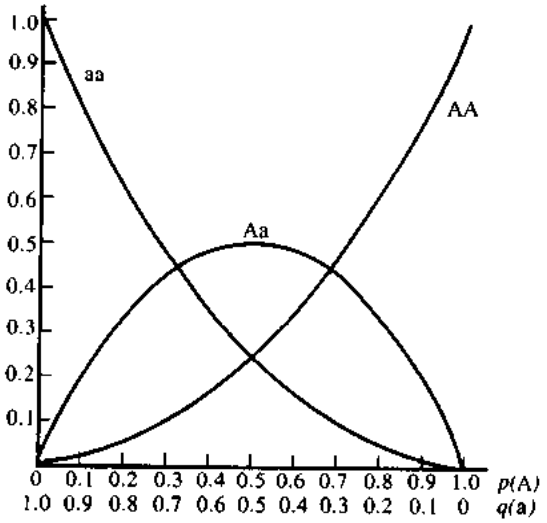


图 25－2 基因型频率基因频率之间的关系。纵座标表示某基因型在平衡群体中所占的比例。横座标表示基因频率。(引自 Griffiths A.J.F. *et al.*: An Introduction to Genetic Analysis, 5th ed., 1993, Fig. 25－10)

四、哈迪－温伯格定律的扩展

当一座位上有两个等位基因时，哈迪－温伯格定律告诉我们在平衡时基因型频率将是 p^2，$2pq$ 和 q^2，它等于等位基因的频率 $(p+q)^2$，假如存在 3 个等位基因 A，B 和 C 的话，它们的频率分别为 p、q、r，在平衡时基因型的频率也等于等位基因频率的平方：

$$(p+q+r)^2 = p^2+q^2+r^2+2qp+2pr+2qr$$
$$= p^2(\mathrm{AA})+q^2(\mathrm{BB})+r^2(\mathrm{CC})+2pq(\mathrm{AB})+2pr(\mathrm{AC})+2qr(\mathrm{BC})$$

沿着北美的大西洋海岸发现兰色的贻贝中有一基因座位编码亮氨酸肽酶，这个座位有 3 个等位基因。随着地理位置的改变这些等位基因频率也发生改变。R.K.Koehn 及其同事确定了栖居于长岛海峡贻贝的群体。

等位基因　　频率

LAP^{96}　$p = 0.52$

LAP^{96}　$q = 0.31$

LAP^{94}　$r = 0.17$

若这个群体达到哈迪－温伯格平衡，预期它们的基因型频率将是

基因型	预期频率
LAP^{98}/LAP^{98}	$p^2 = (0.52)^2 = 0.27$
LAP^{98}/LAP^{96}	$2pq = 2(0.52)(0.31) = 0.32$
LAP^{96}/LAP^{96}	$q^2 = (0.31)^2 = 0.10$
LAP^{96}/LAP^{94}	$2qr = 2(0.31)(0.17) = 0.10$
LAP^{94}/LAP^{98}	$2qr = 2(0.52)(0.17) = 0.18$
LAP^{94}/LAP^{94}	$r^2 = (0.17)^2 = 0.03$

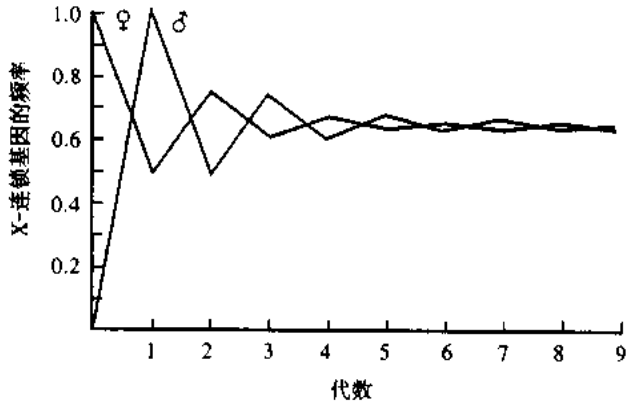


图 25－3　表示 X－连锁基因的频率虽开始时在男、女中差异很大（设开始时男子的 X 连锁基因的频率为 0，女子为 1），但经随机婚配几代后便可达到平衡。

基因频率的平方可以相同的方法来估计一个座位上有更多等位基因的基因型的频率。

若等位基因是 X－连锁的，雌性可以是纯合的或杂合的，但雄性只带有一单个的 X－连锁的等位基因。在雌性中 X－连锁位基因其哈迪－温伯格频率是和常染色体座

位是相同的：p^2（X^AY^A），$2pq$（X^AY^B），和 q^2（X^BY^B）。在雌性中基因型的频率将为 p（X^AY）和 q（X^BY）。和群体中等位基因的频相同。由于这个原因，隐性 X－连锁性状的频率雄性比雌性高得多。为了说明这样一个概念。现以 X－连锁的红绿色盲为例，在不同血统的人类中色盲等位基因的频率如下：美国黑人是 0.039，达到平衡时其色盲男性的预期频率是 $q = 0.039$，而色盲女人的频率 $q^2 = (0.039)2 = 0.0015$。

当在群体中随机婚配时一代以后基因型频率达到平衡。如果等位基因是 X－连锁的，而且在不同性别频率不同。但平衡时基因频率在各代中相似。这是由于雄性仅从他们的母亲那儿接受 X 染色体。当雌性从父母那儿各接受一个 X 染色体。结果雄性中一个 X－连锁等位基因的频率将和他母体体内此等位基因的频率相同。反之雌性中此频率将为其父母中频率的平均数。若随机交配的话，两性中的等位基因频率在每代中来回摆动（图 25－3）。一旦雌、雄的等位基因相等，基因频率在随机交配的几代以后达到哈迪－温伯格比例。

五、哈迪－温伯格比例的测定

为了决定一个群体的基因型是否符合哈迪－温伯格比，我们首先从基因型频率中计算 p 和 q。一旦我们求得了这些基因频率就能计算基因型频率（p^2, $2pq$ 和 q^2），并将这些频率和实际获得和基因型的频率相比较，用卡方来测定。卡方测定使我们有可能掌握我们获得的和用哈迪－温伯格定律计算出的理论频率之间的差异是否是随机的。

为了说明这一过程现以红－黑田鼠（*Clethrionomys gapperi*）的血红蛋白基因为例。人们在这个基因座位发现了 3 种基因型：MM、MJ 和 JJ。1976 年在加拿大西北部发现了一个田鼠群体，群体中有 12 只 MM 型，53 只 MJ 型，和 12 只 JJ 型。为了确定它们的基因型是否符合哈迪－温伯格比，我们先用我们熟悉的公式来计算它们的基因频率：

$$p = \frac{2\times\text{纯合子数}+\text{杂合子数}}{2\times\text{总的个体数}} = f(\mathrm{M}) = \frac{(2\times 12)+53}{2\times 77} = 0.50$$

$q = f(\mathrm{J}) = 1-0.50 = 0.50$

用从基因型中计算出的 p 和 q 能重新计算基因型的哈迪－温伯格比：

$f(\mathrm{MM}) = p^2 = (0.50)^2 = 0.25$;

$f(\mathrm{MJ}) = 2pq = 2(0.50)(0.50) = 0.50$

$f(\mathrm{JJ}) = q^2 = (0.50)^2 = 0.25$。

再用卡方测定来进行检验。为了获得预期理论数我们可将预期的比例乘以实际的总数（N）。（表 25－4），查卡方表，当自由度为 1 时，卡方值为 10.98 时，p 值小于

表 25－4 红-黑田鼠的血红蛋白基因型比的卡方测定

	MM	MJ	JJ
预期值 e	$p^2\times N = 0.25\times 77 = 19.3$	$2pq\times N = 0.50\times 77 = 38.5$	$q^2\times N = 0.25\times 77 = 19.3$
实际值	12	53	12
差数 d	7.3	14.5	7.3
d^2	53.29	210.25	53.29
d^2/e	2.76	5.46	2.76
$\chi^2 = \Sigma d^2/e$		10.98	

0.05，表明观察值与预期值之间有明显差异。也就是说获得的实际数并不符合哈迪－温伯格定律。

六、用哈迪－温伯格定律来估计基因频率

哈迪－温伯格定律的一个重要的应用是将一个或多个等位基因同时用来计算基因频率。例如已知人类的白化是由常染色体隐性基因引起的。一般情况下这一性状是罕见的，但在美国西南部的亚得桑那州的 Hopi 印第安人中白化较多。C.M.Woolf 等在 1969 年访问了 Hopi 村庄，发现总群体约为 6000 人的 Hopi 人中有 26 例白化人，白化频率为 26/6000 或为 0.0043，远远高于绝大部分群本中白化频率。虽然我们可以计算这个性状的频率，但我们不能直接决定白化基因的频率。这是因为我们不能区分杂合的携带者和正常等位基因的纯合体，回顾一下有关等位基因数的基因频率的计算：

$$p = \frac{2\times 纯合子数 + 杂合子数}{2N}$$

但由于无法确定纯合体和杂合体，因此不能用此公式。但如果这个群体达到了平衡的话，我们可以用哈迪温伯格定律来计算。在平衡时，隐性纯合体的频率是 q^2，对于 Hopi 白化来说 $q^2 = 0.0043$。因此 $q = 0.063$，而 $p = 1 - q = 0.935$，那么群体中杂合子的频率是 $2pq = 2\times 0.935\times 0.065 = 0.122$。表明平均 1/8 的 Hopis 人都携带一个白化等位基因。

我们应牢记用哈迪－温伯格定律来计算基因频率的方法。若条件不符合哈迪－温伯格定律，那么我们计算出的基因频率是不准确的。一旦我用此假设来计算基因频率就不能再测定这个群体的基因频率是否适合预期的哈迪－温伯格比。

第三节　在自然群体中的遗传变异

在群体遗传学中有意义的问题之一是要了解自然群体中存在多少遗传变异。群体中可遗传的变异的总数是很重要的，首先它决定了进化变异和适应的潜能。变异的总数也为我们提供了关于各种进化力量相对重要性的思路，因为有些力量增加了变异，而另一些力量减少了变异。新种产生的一种方式可能取决隐含在群体中的遗传变异的总量。由于这些原因群体遗传学对量度遗传变异和控制它们的进化力量感兴趣。

一、遗传变异的模型

在 1940 年到 1950 年这段时期里群体遗传学关于自然群体中遗传变异的总量，形成两种相反的学说，这些学说或遗传变异的模型称为**平衡模型**(balance model)和**经典模型**(classical model)。经典模型是由实验遗传学家提出的，他们认为大部分自然群体具有较小的遗传变异。根据这个模型，在每个群体中一个等位基因的功能是最好的，这个等位基因对自然选择能很好地适应。自然选择的结果使这个群体中的每一个体几乎都是“最好”的等位基因或“野生型”基因的纯合体。新的等位基因逐步通过突变而产生。但几乎所有新的基因都是有害的。由于选择新基因只能以很低的频率存在，偶尔一个新的突变优于野生型基因。这些新的等位基因使个体增加了生存和繁殖的能力，由于它的选择优势，这种等位基因的频率逐代增加。结果新的等位基因达到了很高的频率而成为新的野生型。

群体以这种方式进化,但在任何时期在群体中发生遗传变异总是很少的。

平衡模型预言:没有单个的等位基因是最好的或分布广泛的。一个群体的基因库在每个座位含有很多的等位基因。因此群体中的个体在许多座位上都是杂合的。进化是通过逐步地改变等位基因的频率和种类而发生的。为了说明变异的过程。"平衡模型"论者提出:在一个群体中通过平衡选择,自然选择积极地维持了遗传变异。平衡选择是在等位基因中维持平衡的一种选择,防止任何单个基因达到高频率。平衡选择的一种形式就是**超显性**,在超显性中杂合体比纯合的适应性更高。如在非洲镰刀形贫血的发病率高于其他地方,那是由于镰刀形贫血(Hb^AHb^S)杂合子的红细胞比镰形贫血患者(Hb^SHb^S纯合体)及正常人(Hb^AHb^A)的红细胞对疟原虫具有更高的抵抗力,3种基因型的适应性分别为 Hb^AHb^A:0.81, Hb^AHb^S:1.00, Hb^SHb^S:0.20。这种遗传变异平衡观点是由群体遗传学家提出的,它们有自然史的知识背景和研究野生群体的实践。

二、用蛋白电泳来量度遗传变异

群体遗传学家花了很多年的时间都未能搞清在自然群体中存在多少变异。由于没有一种普遍性程序可以用来估价自然中遗传变异的总量,因此要确定某种遗传变异模型是否正确是很困难的。为了区别经典模型和平衡模型,群体遗传学家需要获得很多物种多个个体的许多座位上基因型的资料。

直到1966都没有一种技术能帮助获得这些信息。博物学家们认为自然界中的动植物的表型常常是不同的,但大部分性状的遗传基础太复杂了,以至难以确定个体的基因型。仅少数的物种具有简单的遗传性状,例如蝴蝶斑点的样式,蜗牛壳的颜色,但这些互不相干的资料太少,而无法进行遗传变异的计算。间接的方法表明很多群体具有大量的遗传变异,但不能直接地去量度。

1966年群体遗传学家开始应用蛋白质电泳的原理来研究自然群体。电泳是用来分离不同结构蛋白质的生化技术,对于群体遗传学来说电泳提供了快速区别很多个体在许多座位上的基因型这一技术。这一程序现在被用于数百种动、植物遗传变异的检测。在一个群体中遗传变异的总量用两个参数来量度,那就是**多态座位比**和**杂合性**(proportion of polymorphic loci and heterozygosity)。

多态座位是在群体中存在一个以上等位基因的座位。多态座位的比(P)是用确定多态座位数目除以被检测座位的总数。例如我们在绿蛙的一个群体中检测了33个座位发现有18个多态座位,那么多态座位的比是18/33=0.55。杂合性(H)是个体杂合座位的比。在绿蛙 的例子中假设在来自一个群体的很多个体的酯酶基因座位上杂合的频率为0.09,也就是说这个座位的杂合性为0.09。我们应将带有此酶的其他座位进行平均,从而获得这个群体的杂合性的值。

表25-5列出了许多物种用电泳方法所获得的多态座位比例如杂合性的值。这些研究的结果很清楚表明大部分物种其蛋白质具有大量的遗传变异,而经典模型明显是错误的。实际上标准的电泳技术丢失了遗传变异的很大比例,因为导致蛋白质在胶上迁移率的改变的变异仅是总变异的一小部分。因此遗传变异真实的量要比通过电泳所揭示的量要大。

在自然群体中蛋白泳动率的变化的研究表明经典模型是不正确的,但并不能证明平衡模型是正确的。平衡模型预言在自然群体中存在很多的遗传变异,但平衡模型也提出大量的遗传变异是通过平衡选择来维持,而电泳研究的结果并不能证实这一论

表 25－5　一些动植物的遗传变异

种　类	种或类型数	每物种座位的平均数	座位的平均比例	
			每个群体的多态座位	每个个体的杂合性
昆虫：				
果蝇	28	24	0.529 ± 0.030	0.529 ± 0.030
二倍体黄蜂	6	15	0.243 ± 0.039	0.062 ± 0.007
其他昆虫	4	18	0.531	0.151
海中无脊椎动物：	9	26	0.587 ± 0.084	0.147 ± 0.019
蜗牛：				
陆地的	5	18	0.437	0.150
海中的	5	17	0.175	0.083
鱼类	14	21	0.306 ± 0.047	0.078 ± 0.012
两栖类	11	22	0.336 ± 0.034	0.082 ± 0.008
爬行类	9	21	0.231 ± 0.032	0.047 ± 0.008
鸟类	4	19	0.145	0.042
啮齿类	26	26	0.202 ± 0.015	0.054 ± 0.005
大型哺乳类[a]	4	40	0.233	0.037
植物[b]	8	8	0.464 ± 0.064	0.170 ± 0.031

a. 人类、黑猩猩、猪尾猕猴、和南部象海豹；b. 主要是异型杂交种。

摘自 Selander, R. K. Genetic Variation in Natural Population. in: F. J. Ayala. Molecular Evolution. Sinauer, Sunderland, Mass.

点。

在这一点上争论的性质发生变化。原来中心的问题是遗传变异存在的量，现在争论的焦点转移到获得的变异是靠什么维持的。一个新的假设取代了经典模型，称为**中性突变**(neutral mutation)学说，它是由日本的木村资生(1968 年)提出的，他承认在蛋白质中广泛地存在着遗传变异，但这些变异对于自然选择是中性的。这并不意味着通过电泳检测的这些蛋白没有功能，而是不同基因型的这些蛋白生理功能是等价的。此自然选择并不作用中性的等位基因，突变和遗传漂变这些随机的过程显示了我们在自然群体所见到的遗传变异的式样。中性突变学说提出在某些位点上的变异也影响适应性，自然选择将排除这些座位上的变异。

区别平衡模型和中性突变学说是困难的，群体遗传学家仍在争论二者的相对的“功过”，现在尚未有完全一致的结论，认为哪种遗传变异模型是正确的。有可能两个模型都有部分是正确的；在某些座位的遗传变异本质上是中性的，而另一些座位上的变异可能受自然选择的作用。

三、用 RFLPs 和 DNA 测序来计算遗传变异的多态性

分子遗传学的技术为检测 DNA 中的遗传变异提供了直接的检测方法，这些方法可以明确地确定在自然群体中存在的遗传变异的量。一种技术是应用限制性内切酶来测定遗传变异。限制性酶可以在 DNA 的特殊顺序上切开双链。大部分限制性活酶识别 4～6 个碱基对，如制性酶 *Bam*H Ⅰ识别的顺序是，$\begin{matrix}GGATCC\\CCTAGG\end{matrix}$ 这个顺序无论在 DNA 的任何地方，*Bam*H Ⅰ都可识别、剪切。限制性酶切片段可以通过琼糖凝胶电泳来进行分离，再用 EB(溴化乙锭) 来染色加以区别，也能通过特异基因探的杂交来识别特定片

段。

假设两个个体在特定 DNA 顺序上有一个或多个核苷酸的差异，而这个差异发生在限制酶识别位点上(图 25－4)。一个个体有一个 DNA 分子带有限制性位点，但另一个体没有此位点，因此 DNA 顺序存在差异。若将这两个个体的 DNA 和限制性酶混合，再经凝胶电泳分离，那么两个个体产生不同的片段样式(如图 25－5 所示)。胶上不同的样式叫做**限制性片段长度多态性**(restriction fragment length polymorphism，RFLPs)。它们表明两个个体的 DNA 顺序存在差异。RFLPs 的遗传是和其他编码性状的等位基因遗传方式相同。只是 RFLPs 不产生任何相应的外在表型而已。它们的表型是用限制性酶切其 DNA 后，在胶上产生的片段样式。RFLPs 也能用作基因作图的遗传标记。它们能提供不同个体中有关 DNA 序列的信息。这些差异仅涉及到 DNA 的小部分，特别是由限制性酶识别的少数核苷酸。然而我们可假设限制性位点在 DNA 上随机发生，因为这一位点并不是作为一个性状存在的。限制性位点存在与否能用来测定序列中的差异。

为了说明用 RFLPs 来计算遗传变异，我们假设从 5 个野生型小鼠中分离 DNA，再用 Southern 吸印将 DNA 转移到硝酸纤维薄膜或尼龙膜上，用探针进行分子杂交，检测 β－珠蛋白基因。典型的一套限制性模式可从图 25－5 中获得。大家都知道每个小鼠都带有两个拷贝的 β－珠蛋白基因，位于一对同源染色体上。这样一个小鼠可能为 +/+(两条同源染色体上都有限制性位点)，+/－(只有一条染色体上有限制位点)，和 －/－(一对同源染色体上都没有限制性位点)(图 25－5)。也就是说 5 只小鼠共 10 条染色体存在 β－珠蛋白顺序，其中 4 个有限制性位点，6 个无限制性位点。为了计算此核苷酸顺序中预期的杂合性，我们可用公式：

$$\text{Hnuc} = \frac{n\left(\sum C_i - \sum C_i^2\right)}{j\left(\sum C_i\right)(n-1)}$$

在此等式中 n 为被测定的同源的 DNA 分子数。在我们的例子中，列了 5 只小鼠，每只有2条同源染色体，因此 $n = 10$。参数 j 表示限制性位点中的核苷酸数。在我们的例子中

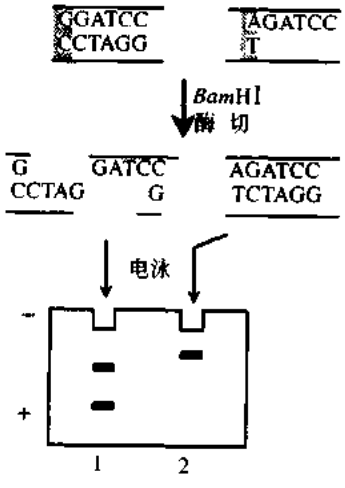


图 25－4 个体 1 和 2 在某 DNA 片段上 *Bam*H Ⅰ酶切位点上仅相差 1 个 bp，经 *Bam*H Ⅰ酶切后电泳可区别。

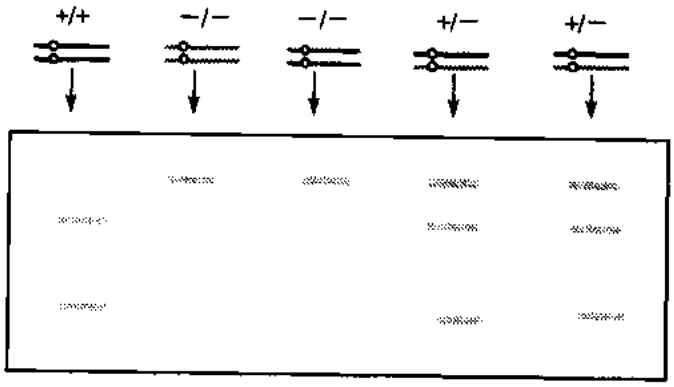


图 25－5 五只小鼠的限制性模式。(+)表示存在限制性位点；(－)表示无限制性位点。

*Bam*H Ⅰ识别 6 个碱基，故 $j = 6$。对于每个限制性位点 i 来说 C_i 表明被酶切的分子数。在我们的例子中只有单个的酶切位点，共 4 个，故 $C_i = 4$。$\sum C_i$ 是在所有染色体中切割位点的切割数，在我们的例子中 $\sum C_i = 4$，这样代入公式

$$Hnuc = \frac{10(4) - (4)^2}{6(4)\ (10-1)} = \frac{40-16}{24(9)} = \frac{24}{216} = 0.11$$

用这个等式设想从单个核苷酸差异产生了 RFLP。

表 25－6　DNA 序列核苷酸异合性的计算

DNA 序列	生物	*Hnuc*
β 珠蛋白基因	人	0.002
生长激素基因	人	0.002
乙醇脱氢酶	果蝇	0.006
线粒体 DNA	人	0.004
H4 基因区域	海胆	0.019

引自 P.J. Russell: Genetics, 3rd. ed., 1992, Table 23.4

通过限制性酶对很多不同区域的核苷酸异合性进行了研究。表 25－6 中举了几个例子。真核生物核苷酸异合性典型的变异从 0.002 到 0.02。这意味着一个个体约有 1/50 到 1/500 的核苷是异合的（在两个同源染色体上含有不同的核苷。）另一种解释核苷杂合性的方法是从群体中随机选择两个染色体，它们之间将有 1/50～1/500 核苷酸有差异。

对遗传变异用 RFLPs 有不利之处，即这种方法仅在基因核苷的很小亚群中揭示了差异。用 RFLPs 分析只用很小的核苷标本（通过限制性酶识别的核苷）来计算变异的所有水平。DNA 测序的方法可以测定一套 DNA 分子中核苷酸全部的差异。例如 M Kreitman 对黑腹果蝇乙醇脱氢酶基因 2659bp 序列的 11 个拷贝（从不同的果蝇中获得）进行了测序，在这 11 个拷贝中，他发现 2659bp 的片段中有 43 位点核苷酸不同。有 8 个不同的等位基因。这表明在群体中 DNA 序列包含着数目巨大的遗传变异。

第四节　群体中基因频率的改变

前面已讲到。当一个群体是大的，随机交配，不受进化力量的影响时，群体的基因频率并不因为遗传而改变。然而对很多群体而言并不具备哈迪－温伯格所要求的条件。如群体常常并不大，交配也不是随机的，另外进化的压力也是存在的。在这种情况下，基因频率会发生改变，而群体的基因库对不同进化因子相互作用作出反应而发生进化。下面我们来讨论在四种进化因子的作用下（突变，基因漂变，迁移和自然选择）怎样改变群体中的基因频率。

一、突变

有一种潜在的能改变群体中等位基因频率的力量那就是突变。正如我们在 21 章所讨论的那样，基因突变在基因的 DNA 中发生了可遗传的改变。突变率一般是很低的，但在物种中各座位之间的突变率是不同的，有些基因影响到所有的突变率，还有很多环境因素，如化学物质、放射线、传染介质等可能增加突变率。

进化有二个阶段：第一是遗传变异的产生；二是不同等位基因频率的增加或减少来对进化力量作出反应。突变在两个阶段中具有潜在的重要性。

最后，突变是一切新遗传变异的来源，等位基因的新组合可以通过重组而产生，但

新的等位基因只有当突变发生时才能产生。突变提供了进化作用的遗传原料。大部分的突变将使个体受到损伤，并从群体中被排除。然而少数的突变可能对个体是有利的，且在群体中传播开。突变有害还是有利取决于特殊的环境，若环境改变了，先前有害的突变可以变成有利的突变。例如在使用杀虫剂 DDT 之后，昆虫产生了对 DDT 的抗性突变，这样昆虫在喷有 DDT 的环境中不仅可以生存还能繁殖。由于这种优势，对 DDT 抗性突变在昆虫的群体中迅速地得到发展。突变是进化过程中重要的因素。因为它们提供了遗传变异，这正是进化力量所作用的对象。

突变对进化还有一种潜在的作用，这是通过进化过程的第二步实现的，此就是通过改变群体中等位基因频率。例如一个群体由 50 个个体组成，若它们的基因型都是 AA，那么 A(p)的频率是 1.00 [(2×50)/100]。若 1 个"A"等位基因突变成了"a"，那么这个群体变成由 49 个"AA"和 1 个"Aa"组合组成。那么现在 p 的频率是[(2×49)+1]/100=0.99。当再发生一次这样的突变 A 的频率降到 0.98。若这些突变持续发生下去，经过很长的时间最终 A 的频率将降到 0，而 a 的频率将升为 1.00。

突变过程对等位基因的影响是和下列情况相似：如果我们有一大罐黑色的围棋子，每年我们取出一粒而放进去一粒白子，那么最终这围棋子全是白的。与此相似，突变反复发生虽然很慢，但只要维持下去就可能导致基因频率的改变，当然这是在假定不存在其他进化力量的情况下。

A 突变为 a 称为**正向突变**(forward mutation)。大部分基因也会发生反向突变，即 a 突变 A，这些突变称为**回复突变**(reverse mutation)。在一般情况下正向突变的频率要大于回复突变的频率。为了说明回复突变，假设开始我们用一罐黑棋子，现在不是每次用白子去替换黑子，而是随机的取一粒白子或黑子来取代。在这个样本系统中不存在对棋子黑、白颜色的选择。开始时似乎是总有黑子被白子取代，但到罐中的白子比较多的时候，就有可能取出的是白子，加进去的是黑子。在达到平衡时黑子和白子的频率相等。在此时就难以进一步改变白子和黑子的频率。

在围棋子的例子中正向突变和回复突变的比率是相等的。而实际上正向突变和回复突变的频率是不同的。设正向突变的频率(A→a)为 u，回复突变(a→A)的频率为 v。群体中 $f(A)=p$，$f(a)=q$，假设群体很大，无自然选择存在，那么在每一代中，A 等位基因以 u 频率突变成 a。实际突变数取决于 u 和 A 等位基因的频率 p。即 pu 或($1-q$)u。例如设一个群体含 100 000 个等位基因，若 $u=10^{-4}$，即每 10 000 个 A 等位基因就有一个突变为 a。当 $p=1.00$ 时，群体中所有等位基因(100 000)全部为 A，那么将有 $10^{-4}\times 100\ 000=10$ 个 A 突变为 a，但如果 $p=0.10$，那么仅有 $10^{-4}\times 10\ 000=1$ 个 A 基因突变为 a。由于突变 A→a，A 的频率减少为 up。当回复突变发生(A←a)时，A 等位基因的频率将增加，值为 vq。因此在一代中 A 基因的频率由于正突变而减少，由于回复突变而增加，当正突变和回复突变同时存在时，A 等位基因频率的改变为 $\Delta p = qv - up$。我们可以从公式看出，当 p 值高时，A 等位基因突变为 a 的数相对较大，但 A 等位基因突变为 a 越多，p 值随之减少。在同一时间中，当 q 很小时，回复突变的等位基因很小。当 q 增加时，回复突变的等位基因增加。最终这个群体达到平衡，也就是 A 和 a 的频率都保持定。

若 $up > vq$ 时，f(a)增加，即正突变 > 回复突变，

若 $up < vq$ 时，f(A)增加，即正突变 < 回复突变。

若达到平衡时，$\Delta p = vq - up = 0$，$vq = up$，$vq = u(1-q)$

解这个等式，我们可以得到平衡时 q 值(以 q 表示)

$$vq = u - uq$$
$$vq + uq = u$$
$$q(v+u) = u$$

p 的平衡值 $p = 1 - q$，若 $\hat{q} = \dfrac{v}{u+v}$ 用 p 取代 q 解这个等式，得 $\hat{p} = \dfrac{v}{u+v}$ 设一个群体开始时基因频率 $p = 0.9$，$q = 0.1$，正向突变和回复突变的平分别为 $u = 5 \times 10^{-5}$，$v = 2 \times 10^{-5}$，那么第一代基因频率的变化为

$$\Delta p = vq - up = (2 \times 10^{-5} \times 0.1) - (5 \times 10^{-5} \times 0.9) = -0.000043$$

$f(A)$ 的减少仅 0.43‰。如果其他的因素作用此群体，经过多代后可能达到平衡，因此一个群体在没有进化力量作用下，其突变率决定了等位基因频率。但由于突变率如此之低，因此由于突变使基因频率的改变也是非常之低。此外一个基因频率接近平衡值时频率的改变越来越小。设 A→a 的突变率为 $u = 10^{-5}$，而无回复突变的发生，则 A 的平衡频率是 0.0。若 f(A) 从 0.50 变为 0.049 需 2000 代，f(A) 从 0.1 变为 0.09 则需 10 000 代。若存在回复突变，频率的改变甚至更慢。

实际上突变本身改变基因频率是很慢的，群体很少能达到突变的平衡。另外的一些因素如自然选择对基因频率的影响要大得多。突变是惟一很少能决定一个群体的基因频率的因素。例如软骨发育不全的侏儒症是人类常染色体显性性状，虽然这种突变是有重复性，但在人类群体中的频率是由突变和自然选择相互作用而决定的。

二、遗传漂变

哈迪－温伯格定律的一个主要的假定就是群体无限大。实际的群体不可能无限大，但它们要大到足以能预期基因频率和基因型频率，而且一些机率因素可以对基因频率起到重要的作用。但有些群体很小，这些群体的机率因子可使基因频率产生随机变化。由于样本的机误（chance errors）导致群体基因频率的随机改变称为**遗传漂变**（genetic drift）或简称为漂变。此是由群体遗传学家 S. Wright 于 1930 年提出的，有时人们也把漂变称 Wright 效应。

1. 基因频率的随机改变

由于机误而产生的基因频率的随机改变在小的群体中是一种重要的进化的力量。设想在南太平洋岛上群住着人类的最小群体，假设只有 10 个人组成，5 个人绿眼睛，5 个人棕色眼，假设眼睛的颜色是由单基因决定的（实际上眼睛的颜色是由多基因遗传的），且绿色眼的等位基因 b 对棕色眼的等位基因 B 是隐性的（BB 和 Bb 为棕色眼，bb 为绿色眼）。此岛上的群体中绿眼等位基因的频率是 0.6。一次台风袭击了这个岛，群体中 50%的人死亡了，也就是 5 位居民死于风暴。正巧这死去的 5 个人是棕色眼。眼色绝不会影响存活的概率，其实在这次风暴中绿色眼的人都得以幸存完全出于偶然。台风之后这个岛上绿色基因的频率是 1.0。在这个群岛上的群体中所发生的突变是绿眼基因的频率从 0.6 变成了 1.0，这完全是机会所造成。

现在设想这个群体不是 10 个人，而有 1000 人，和前面一样群体中 50%是绿色眼，50%是棕色眼，一次台风袭击有一半人死于灾害，若死亡是随机的，那么是否可能正好

死亡的都是棕色眼呢?在这样一个 1000 人的群体中,随机出现这样情况是概率是极小的,这个例子是为了说明遗传漂变的重要特点:偶然的因素只能在很小的群体中使基因频率发生改变。

在自然群体中偶然产生的死亡率,例如前面例子中的台风仅是产生遗传漂变途经中的一种。与配子和合子预期比例产生的偶然偏差也会导致遗传漂变。我们已知在遗传杂交中和预期比例发生偏差的重要性。例如将一个杂合子和一个纯合子杂交(Aa × aa),预期后代中 Aa : aa 为 1 : 1,但我们不能预期每次的结果都是精确的 1 : 1。若后代的数目比较小,那么获得的实际比例可能和预期值产生较大的偏差。而哈迪 - 温伯格定律是建立在随机交配来预期这种交配所产生的后代的比(表 25 - 3)。后代的实际数与预期数目往往有偏差,那么基因型频率可能就不符合哈迪 - 温伯格比,结果使基因频率发生改变。

机率使结果和预期比产生偏差的这种现象被称为**取样误差**(sampling error)。设群体产生一个无穷大的配子库,其含有的等位基因比分别为 p 和 q。若发生随机交配,所有的配子都结合成合子,那么基因型的比将是: p^2, $2pq$ 和 q^2, 而在合子中等位基因的频率仍保持为 p 和 q。也许后代的数目是有限的,但这些与配子结合形成的后代构成了潜在的无限大配子库中的一个样本。由于随机性,或由于样本的机误,样本可能偏离大库,样本越小,偏差越大。

掷硬币的情况和取样误差相似。当掷硬币时我们预期 50% 是正,50% 是反。若掷了 1000 次的话,获得的结果很接近一半对一半的比。但我们仅掷 4 次的话,若得到 3 正 1 反,或者全都是反的也不足为奇。表明样本越小,取样误差越大。所有的遗传漂变都是由取样误差所造成的。

2. 遗传漂变的量度

遗传漂变是随机的,我们难以预言漂变后基因频率的多少,但取样误差是和群体的大小相关的,所以我们就能预言漂变的大小。生态学家常常通过计算个体的数目来量度群体的大小,但不是所有的个体都能将配子供给下一代。为了确定遗传漂变的大小,我们必须知道有效种群大小(effective population size),等于成体为下代提供的配子的当量数。若两性的数目相等,而每个个体产生后代的概率都相同,那么这个有效群体的大小等于此群体中交配的成体数。然而当雌性数不等时,有效群体的大小是:

$$N_e = \frac{4 \times N_f \times N_m}{N_f + N_m}$$

等式中 N_f 是交配的雌体数, N_m 是交配的雄体数。

为什么要用这个等式计算有效群体的大小不是简单地用交配的成体数呢?原因是雄性作为一个群体为下一代全部基因提供了一半来源。而雌性则提供另一半基因来源。因此 70 个雌性和 2 个雄性的群体中,2 个雄性并不简单地与二雌性相当,每个雄性要为下一代总的基因数贡献 $1/2 \times 1/2 = 0.25$。而每个雌性只贡献所有基因的 $1/2 \times 1/70 = 0.007$。数目很小的雄性不均衡地影响着下一代存在的等位基因,利用上述等式,有效群体的大小是

$$N_e = (4 \times 70 \times 2) / (70 + 2) = 7.8$$

或近似等于 8 个交配成体。这意味着 70 雌和 2 雄的群体其遗传漂变相当于只有

4个交配雄性和4个交配雌性的小群体的遗传漂变。

其他因素如后代的差异，群体大小的波动，重叠的后代能进一步减少有效群体的大小。考虑到很多影响因素，精确计算有效群体的大小是困难的。

由遗传漂变导致群体中总量发生变化可以通过基因频率的**方差**（variance of gene frequency）来计算，等式为：$S_p^2 = \frac{pq}{2N_e}$

公式中 Ne 为有效群体的大小。p 和 q 是基因的频率，常用的计算是基因频率的标准误（standard error of gene frequency），即基因频率方差的方根 $S_p = \sqrt{\frac{pq}{2N_e}}$

此标准误可用来计算 95% 的基因频率置信界限。表明在这个群体中 p 的范围预计在 95%。95% 的**置信界限**（confidence limits）近似等于 $p + 2S_p$。例如在一个群体中 $p = 0.8$，$Ne = 50$，那么基因频率的标准误差是：$S_p = \sqrt{\frac{pq}{2N_e}} = 0.04$。对 p 来说 95% 置信界限是

$$p \pm 2S_p = 0.72 \leqslant p \leqslant 0.88$$

为了解释 95% 置信界限，设 Ne 为 50 的 100 个群体开始时 $p = 0.8$。遗传漂变可以导致某些群体中基因频率发生改变；95% 置信界限告诉我们在下一代中，原来的 100 个群体其 p 的范围在 0.72 到 0.88 之间。因此如果我们观察到 p 的改变大于以上范围，如从 0.8 到 0.68，则我们知道由于遗传漂变将要改变的概率小于 0.05。很有可能我们将考虑的是另一些力量，而不是遗传漂变对基因频率的影响。

3. 遗传漂变原因

所有的遗传漂变都是由于取样误差而产生的，但在自然群体中取样误差的发生有各种途径。首先是当群体大小经很多代都保持很小。毫无疑问这种情况是经常发生的，特别是在栖息地的边缘。或者当竞争限制了群体生长时，在这种群体中遗传漂变在基因频率进化中起到了重要的作用。Buri 用果蝇做了一个实验，表明了遗传漂变的作用。在果蝇决定眼色的座位上有两个等位基因 bw^{75} 和 bw。他建立了 107 个实验群体。bw^{75} 开始时的频率为 0.5，在各群体中果蝇随机互交，然后每代随机地选出 8 雌，8 雄作为下一代的亲体。这样有效群的大小为 16 个果蝇。在 107 个群体中等位基因频的分布如图 25 - 6 所示。注意在早期的世代中基因频率群集在 0.5 周围，但遗传漂变导致这个群体的基因频率分散开，或逐代分开。在大部分群全中到 19 代 bw^{75} 的频率为 0 或1。

遗传漂变产生的另一种途经是通过建立者效应（founder effect）。建立者效应发生在由少量的个体开始建立群体时。虽然群体随后可以增大，以后可有大量的个体组成，但群体的基因库源出于最初建立时存在的基因。机率在建立时确定存在什么基因起到了重要的作用。此对于后来多代的基因库有着深远的影响。许多经典的建立者效应实验来自人类群体的研究。在大西洋南部有三个很小的火山岛（Tristan da cunha），1817 年开始由苏格兰人 William Glass 和他的家族到岛上居住。他们又迁来几个失事的水手和几个来自远处岛上的妇女，但这个岛是保持着遗传隔离的状态。1961 年岛上火山爆发，岛上的群体几乎有 300 个居住者组成的群体到英格兰。他们在英国呆了 2 年，遗传学家研究了他们，并重建了这个群体史。这些研究者们揭示了这个群体当年的基因库，

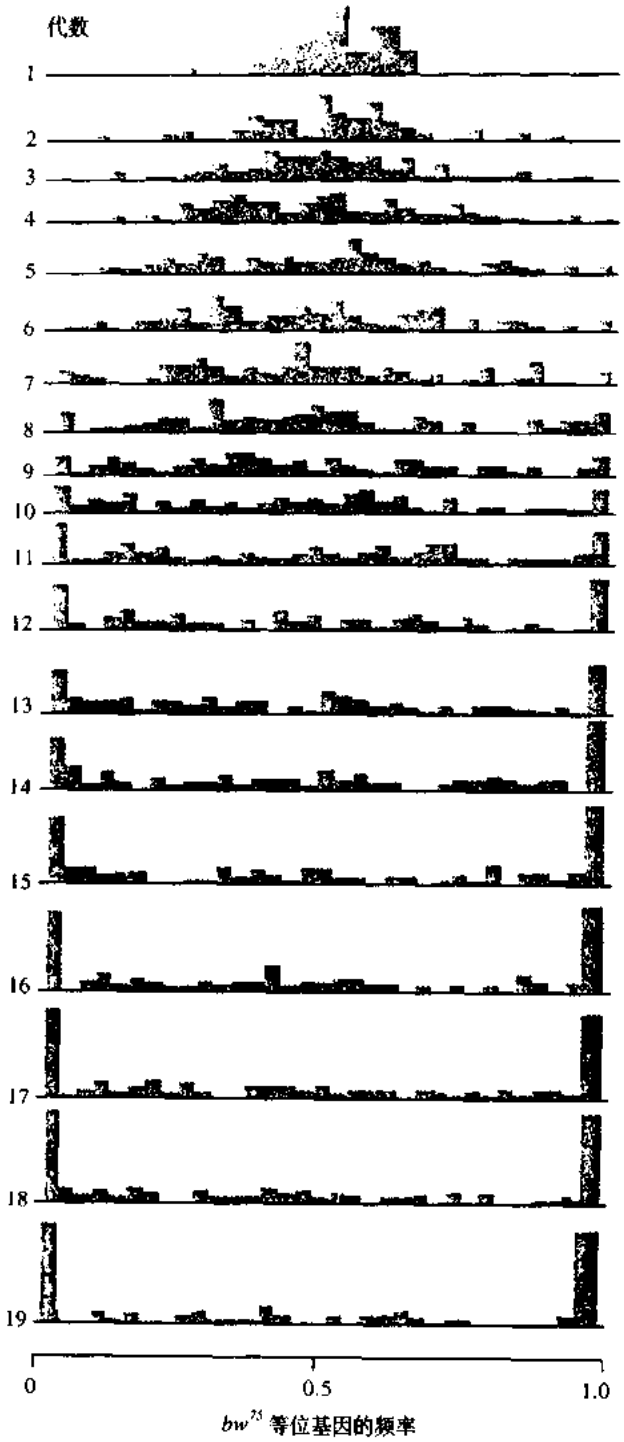


图 25－6 Buri 对 107 个果蝇群体的遗传漂变研究的结果。图显示了这些群体连续 19 代等位基因频率的分布，每个群体有 16 只果蝇，雌雄各半。（引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 23－10）

其基因频率强烈地受到了遗传漂变的影响。

岛上的群体在突变中发生了三种类型的遗传漂变。① 建立者效应发生在开始定居者中。到 1855 年这个岛的群体扩大到约 100 人，但这个群体 26%的基因是由 William Glass 和他妻子传下来的。甚至到了 1961 年这 300 人的群体的全部基因中有 14%的基

因来自最初的两个定居者。Glass 和其他的原来的建立者们的特殊基因对以后群体的基因库有着重大的影响。② 群体一直保持很小，取样误差也持续发生③取样误差的第三种形式称为**瓶颈效应**(bottleneck effect)这种效应在上面所提到的小群体中也起到了重要的作用。当一个群体减少得很厉害时发生瓶颈效应。在这样的随机群体中，偶然事件可能使某些基因从基因库中丢失。我们前面举的 10 个人小群体眼色的例子就是如此。当飓风袭击这个岛时，从而使绿眼的基因频率从 0.6 变成 1.0。这就是瓶颈效应的例子。瓶颈效应也可以看成是建立者效应的一种类型，因为仅几个个体的减少结果就会影响群体。

在 Tristan da Cunha 岛的历史上曾发生了两次剧烈的瓶颈效应。第一次约在 1856 年突然发生两件事：那是 William Glass 的死和一个传教士的到来，他鼓励当地居民离开这个岛。当时很多的岛民移居到美国和南非。这个群体在 1855 年末还有 103 人，但到了 1857 年只剩下了 33 人。第二次瓶颈效应是发生在 1855 年。Tristan da Cunha 岛没有天然港、岛民们是划着小船到海上的商船上进行贸易。在 1855 年 11 月 28 日这一天，岛上 15 名男性成年人又划船出海了，在全体岛民的目光下，他们离开了岛，但不幸的是船翻了，15 人全部沉入海底。遇难之后岛上，很多寡妇和其子女都在几年内先后离开了这个岛，只余下 4 名成年男人，二名老太太和一名神经错乱的女人。这个群体从 106 人一下又锐降到 59 人。两次的瓶颈效应对这个群体的基因库起到了重要的作用。来自几个岛民的基因都全部丢失了。通过两次事件其他一些先民所传下的基因的相关性也发生了改变。这样 Tristan da Cunha 岛的基因库受到遗传漂变的影响，这些影响是以三种形式存在的，①即建立者效应；②小群体大小；③瓶颈效应。

后面我们将要讨论迁移，在群体中**基因流**(gene flow)增加了有效群体的大小而减少了遗传漂变的影响。小的交配单位丢失了基因流，最初是从另一群体中分离出来的，而且常经历了相当大的遗传漂变，甚至受到很大群体的包围。一个很好的例子是一个宗教的教派浸礼会叫做 Dunkers，它在美国东部宾夕凡尼亚。在 1719 年到 1729 年之间，有 50 个 Dunker 家庭从德国迁到美国定居，从此 Dunkers 保持了一个隔离的小群体，不能和教派之外的人群通婚，而他们本身的人数又相当少。

1950 年遗传学家研究了一个在富兰克林城的初始 Dunkers 团体，这个群体约有 300 个成员，经过很多代这个群体的大小仍保持相对稳定。研究发现 Dunkers 的基因频率不同于美国一般的群体。也不同于 Dunkers 的本土西德的群体。表 25－7 表明在 ABC 血型座位上的基因频率。Dunkers 中 ABC 等位基因频率既不同于美国的群体，也不同于西德的群体，也不是居于二者的频率之间。若是在二者之间就有可能是 Dunker 和美国人基因混合而发生的。看来 Dunker 特殊的基因频率可能是漂变的结果，即基因库中的基因发生随机的变化当最初的 50 个家庭从德国移出时可能发生了建立者效应。而遗传漂变从 1729 年开始就持续地影响着每一代的基因频率，这虽因为这个群体的大小一直保持很小。

4. 遗传漂变的效应

遗传漂变引起了基因频率的改变，而这些改变对群体的遗传结构有着明显的作用。首先是遗传漂变导致基因频率逐代改变。图 25－7 说明了这个问题。不同的线表示 5 个群体超过 30 代的基因频率。虽然各群体开始等位基因的频率都是 0.5，各群体都由于取样误差而引起了逐代基因频率的改变。在每代中基因频率可能增加或减少，

从而使逐代频率随机波动和漂变。有的基因频率的值达到了 1.0,有的又降到零。即一个等位基因在群体中被固定或丢失。一旦一个等位基因被固定了,其基因频率就不再发生变化。除非另一个等位基因通过突变或迁移被重新引入。在一个群体中固定的概率随时间的推移而增加。如图 25-8 所示。如开始时的基因频率是相等的,等位基因要能成为固定完全是随机的。另一方面如果基因频率开始不等,频率较低的等位基因可能会被丢失。

在基因漂变或固定的过程中,群体的杂合子数也将减少,固定之后群体中杂合体为零。当杂合性减小时等位基因逐渐被固定,群体便失去遗传变异。这样遗传漂变的第二个效应就是减少群体中的遗传变异。

由于遗传漂变引起了基因频率的随机改变,等位基因的频率发生分离。单个群体在相同方向上将不会发生改变。因此通过遗传漂变群体的基因频率会发生歧化。在图 25-7, 8 中表明一个群体开始时一对等位基因的频率 p 和 q 都等于 0.5, 几代之后等位基因频率发生歧化,而这种歧化是随着代数的增加而增加。当所有的群体一对等位基因中一个被固定时其基因频率歧化达到极大值。若基因频率开始时都等于 0.5, 则约有半数群体中有一个等位基因被固定,另一半则固定另一个等位基因。

由于在小群体中遗传漂变很剧烈,因而导致了遗传的歧化。不难予料在小群体中基因频率的变异要大于大群体。在自然群体研究中人们就获得了这样一种相关性。

表 25-7 三种人群中 ABO 血型系统等位基因频率和表型频率

群体	等位基因频率			血型表型的频率			
	I^A	I^B	i	A	B	AB	O
Dunkers	0.38	0.03	0.59	0.593	0.036	0.023	0.348
美国	0.26	0.04	0.70	0.431	0.058	0.021	0.490
西德	0.29	0.07	0.64	0.455	0.095	0.041	0.041

引自 Peter J. Russell: Genetics, 3rd ed., 1992, Table 23.6.

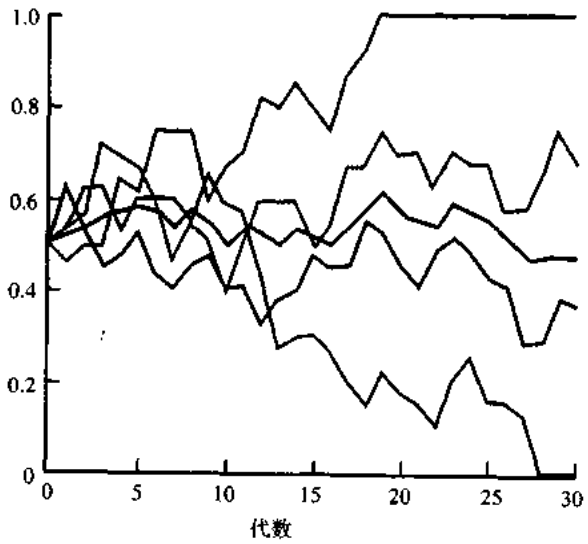


图 25-7 表示 5 个不同群体超过 30 代的基因频率的变化。虽然各群体开始等位基因的频率都是 0.5, 但各群体因大小不同, 基因漂变的程度也不同。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 23-11)

R.K.Selander 研究了栖居于德克萨斯谷仓中家鼠群体的遗传变异。通过系统地捕获，他能估算出群体的大小。并用电泳来测定两个座位的基因频率，一个座位编码酯酶（Est－3），另一座位编码珠蛋白。他发现小群体中基因频率的变化要比大群体大好几倍(表 25－8)。他观察的结果符合我们所了解的遗传漂变如何导致群体的歧化。

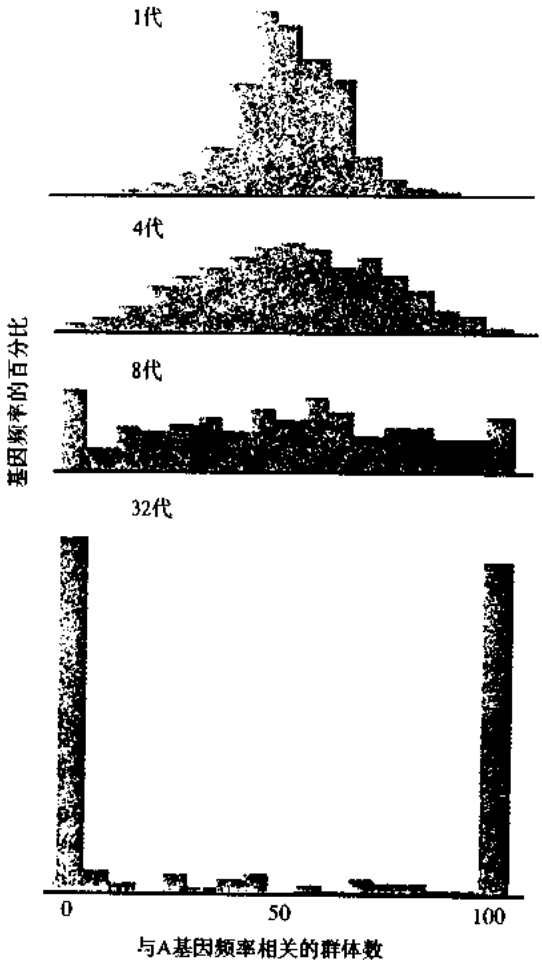


图 25－8　经过遗传漂变等位基因被固定。图表示 400 次实验的结果：每个实验的群体有 8 个二倍体个体，它们的 A 等位基因初始频率都为 0.5。在每个群体中基因频率随着时间而逐步漂变。上面的图表示一代后在 400 个群体中基因频率的分布。中下部的图表示 4，8 和 32 代基因频率的分布。(引自 Russell, P. J.: Genetics, 3rd ed., 1992, Fig. 23－12)

表 25－8　大小不同群体的基因频率和基因频率的变异的比较

群体类型	群体数	等位基因平均频率		等位基因频率变异	
		$Est\text{-}3^b$	Hbb^s	$Est\text{-}3^b$	Hbb^s
小（N＜50）	29	0.418	0.849	0.051	0.188
大（N＞50）	13	0.372	0.843	0.013	0.013

引自 Selander, R. K.: Am Zoologist, 1970, 10:53.

5. 迁移

哈迪－温伯格定律的假设之一是要求群体是封闭而不受外部环境因素的干扰。然而很多群体都不是完全隔离的，它们和相同物种的另一些群体之间交换基因。一些个体迁入另一群体可以使此群体的基因库导入新的基因，而改变原来的等位基因频率。因此**迁移**(migration)具有潜在的打破哈迪－温伯格平衡的作用，从而干扰了群体基因频率的变化。

迁移这个术语通常应用于生物个体从一个群体转入到另一群体。然而在群体遗传学中我们感兴趣的是基因的移动。基因的移动仅发生在生物或配子“迁移”的时候把它们的基因贡献给受纳群体的基因库中。这个过程也可叫做**基因流**（gene flow）。

基因流对群体有二个主要的作用：①它将新的等位基因导入到群体中。由于“突变”这一事件一般是很少发生的，一个特殊的突变等位基因可能在一个群体中发生，而不在另一个群体中发生，基因流可将独特的等位基因播散到其他的群体，像突变一样，成为该群体的遗传变异的来源。②当迁移动物的基因频率和**受纳群体**的不同时，基因流改变了受纳群体的等位基因频率。通过基因的改变，群体保持相似性，这样，迁移是一种均化力量，倾向于阻止群体发生变异。

为了说明迁移对基因频率的作用，现设想一个模型，在模型中基因流仅单向的发生，即从群体Ⅰ移向群体Ⅱ。假设在群体Ⅰ中 A 等位基因频率（P_I）是 0.8，而在群体Ⅱ中 A 的频率（P_{II}）是 0.5。每代某些个体从群体Ⅰ迁移到群体Ⅱ。这些迁移者是群体Ⅰ中基因型随机的样本。迁移后群体Ⅱ实际含有两组个体：一组是迁移者，其 A 等位基因频率 $P_I = 0.8$。另一组是接纳群体的成员，他们的 A 等位基因频率 $P_{II} = 0.5$。迁移者现在在群体Ⅱ中的比例为 m，那么迁移后群体Ⅱ中 A 基因的频率是：

$$P'_{II} = mP_I + (1-m)P_{II}$$

我们可以看到迁移后 A 的频率是由新组成群体Ⅱ中的两组中 A 等位基因的比例来决定。等式右边的第一个成份 mP_I 是迁移者中 A 等位基因的频率。第二个成份是原有者中 A 等位基因的频率 $(1-m)P_{II}$，$(1-m)$ 是受纳群体中原有个体的比例（图 25－9）。这两个成分相加就构成了群体Ⅱ迁移后 A 基因的频率。

在群体Ⅱ中基因频率的改变设为 ΔP，等于混合的 A 频率减去群体Ⅱ原来的 A 基因频率：

$$\Delta p = P'_{II} - P_{II}$$

将上式代入

$$\Delta p = mP_I + (1-m)P_{II} - P_{II}$$

展开

$$\Delta p = mP_I + P_{II} - mP_{II} - P_{II}$$

$$\Delta p = mP_I - mP_{II}$$

$$\Delta p = m(P_I - P_{II})$$

最后结果表明迁移使基因频率发生改变，此依赖于两个因素：混合群体中迁移者的比例(m)和两个群体之间基因频率的差($P_I - P_{II}$)。若迁移所涉及的两个群体之间基因频率无差别，即 $P_I - P_{II} = 0$，那么 Δp 也为零。为了讨论迁移对基因库结构的影响，

必须假定两群体之间基因频率是不同的。由于不断地相互迁移，使 P_I 和 P_{II} 逐渐相同，结果迁移对基因频率没有影响。此时另一个因素 m 对基因频率也不能有所影响了。

移居美国的黑人，他们血型的表现型频率既不同于他们原来的群体非洲黑人，也不同于受纳群体美国的白人，而是倾向于两者之间，这是当地部分白人与黑人通婚的结果，但所有的混血儿都被人们看作为黑人，故白人的频率几乎没有受到影响，只有"黑人"群体的基因频率有了明显的改变，事实上这个混合群体(并未充分混合，因二者之间的婚配不是随机的)总的基因频率也发生了改变(表 25 - 9)。美国黑人血型基因频率的改变和前面所说的浸礼会成员血型频率改变(表 25 - 7)的机理是不同的，前者是漂变，后者是迁移。

表 25 - 9　美国黑人、白人及非洲黑人的血型频率

血型	非洲黑人	加州黑人	加州白人
A	0.156	0.175	0.247
B	0.136	0.125	0.068

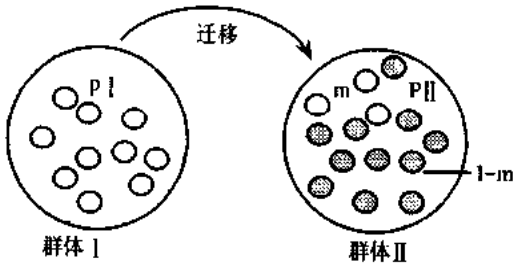


图 25 - 9　迁移后的混合群体。

另外我们要注意群体中的迁移倾向于增加群体的有效大小。正如我们前面已说到的小群体会导致遗传的漂变，而遗传漂变会使群体歧化。在另一方面，迁移会减小群体的歧化，有效增加单个群体的大小。

三、自然选择

我们已经例举了三种能改变基因频率和产生突变的进化力量 - 突变、遗传漂变和迁移。这些力量能改变群体的基因库，而且对物种的进化也有一定的影响。但这 3 种作用都不会产生适应。适应是通过性状的演化使生物体更加适应于环境的过程。这些性状增加了生物生存和繁殖的机会。适应反映了自然界中我们所见到的很多神奇的有趣的性状，如蜂雀的翅膀能使其向后飞。猪笼草的叶片能捕获和吞食昆虫。脑可以使人类思维、说话、阅读和爱恋。这些生物学特点和无数其他的精致的性状都是适应的产物。遗传漂变、突变和迁移都影响到适应的方式和过程，但唯有自然选择才能产生适应。**自然选择**(natural selection)是很多性状的进化的主宰力量，并在自然中形成了很多的表型变异。

达尔文(Charles Darwin)和华莱士(Russel Wallace)在 19 世纪的中期各自独立地建立了"自然选择"这个概念(某些早期的博物学家们已提出相似的想法)。在 1858 年达尔文和华莱士的学说介绍到伦敦的林奈学会，被其他一些科学家热情地接受了。达尔文提出的进化论比华莱士的更为深入。数以百计的观察资料有力地支持了这些学说，《物种起源》这本书的出版使进化论的思想得到了普及。他的卓越贡献使我们了解了自然选择。

自然选择的本质是基因型差别复制，它的含义是：带有某些基因产物的个体比另一些具有更多的后代。这些基因在下一代中的频率得到了增加。通过自然选择对生存

和繁殖有利的性状逐代增加。生物以这种方式来适应它们的环境。

1. 自然群体中的选择

在自然群体中关于选择的一个经典的例子是和工业污染相关的黑色的蛾子的形成，也叫做**工业黑化现象**。在欧洲、北美及英国的工业区中，黑色表型在自然界研究得最多的一种是棕色桦尺蠖（*Biston betularia*），又叫做椒花蛾，这种蛾子的一般表型是灰白色的背景，在身体和翅膀上有细小的黑的斑点。

在 1848 年以前捕获到的椒花蛾都是灰白色的，但到 1848 年在英国工业城市曼彻斯特附近捉到了一只黑色的椒花蛾。这种新的表型称为**黑化型**（*carbonaria*）。它是由突变而产生的。但在曼彻斯特周围和另一工业区中其比例迅速上升，到 1880 年黑蛾的比例已超过了 90%。黑蛾的高频率是和工业区相关联的，在那儿灰白色表型的蛾子都已很少发现。很多学者包括 E.B.Ford 和 R.Goldschmidt 对这一现象进行了研究。表明黑化型对灰白色表型是显性的。而且还发现了第三种表型是介于黑化型和灰白色型之间。叫做 *insularia* 这是由不同座位上一个显性等位基因产生的。

H.B.D.Kettlewell 研究花椒蛾颜色的多态性表明黑化型的增加是在被污染的树木上对灰白色型有很强的选择性所致。椒花蛾是在夜间活动。白天停歇在树干上。鸟类常常在白天捕食这些蛾子。由于黑色蛾在被煤灰弄黑了的树干上形成了保护色，不易被发现，而灰白色的蛾在黑色的树干上颜色的反差很大，成了鸟类捕捉的明显目标。而与此相反白灰色的蛾子在布满苔藓的未被污染的树干上不易被发现，而黑色的突变型在此背景上变得很醒目，易被鸟类捕食（图 25－10）。在工业区，工业污染是从 19 世纪中期开始，树干上的苔藓和地衣大多被杀死了，树干覆盖着黑色煤灰。由于黑色蛾在黑色树干上生存率大大高于灰色蛾，因此有更多的机会将基因传给下一代。这样在工业区里黑色蛾的表型频率增加，灰色表型频率下降，相反在非工业污染区黑色蛾仍十分罕见。

在 20 世纪 50 年代中 Kettlewell 进行了一系列的实验，证实了影响蛾类表型频率变化的主要因素是工业污染和鸟类选择性捕食。他先在林区放出活蛾并观察其结果，证

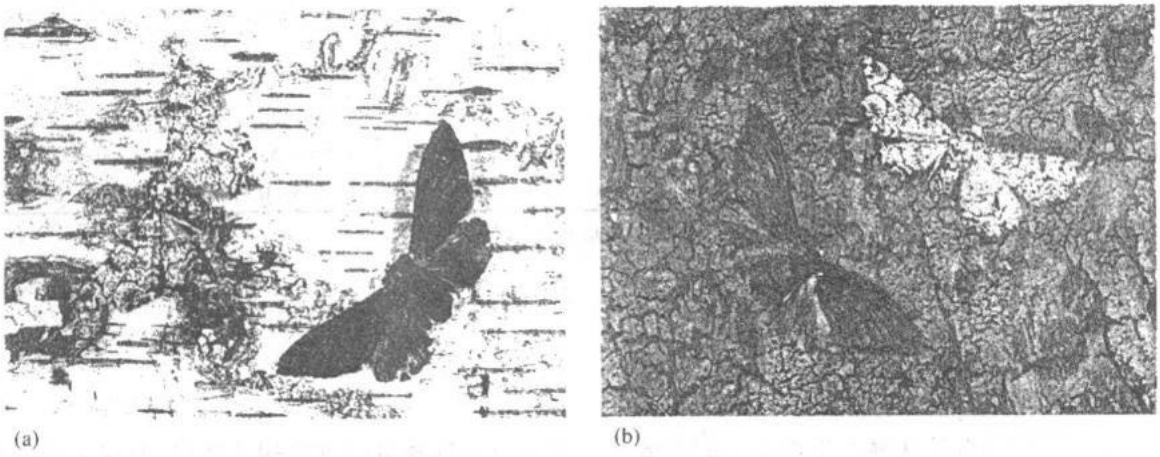


图 25－10　灰色桦尺蠖（*Biston betularia*）和黑色的突变型。(a)在未污染的树干上灰色蛾和背景的颜色相近，不易被发现。(b)而黑色蛾确很明显。在被污染的黑色树干上情况正相反。（转引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig. 23－17）

实那些最醒目的类型被鸟类吃掉的比例最高。然后他在约明翰一个被工业煤烟薰黑的林子里重复做释放和再捕捉实验，还在多塞特一个未被污染的林子里做对照实验结果都显示具有保护色的类型存活比例较高。

20 世纪 50 年代后英国的大多数工业区注重了对环境污染的控制，与此同时蛾类不同颜色的表型比例也出现了反向变迁。如 1959 年黑色型为 93.3%，而到 1965 年就下降到 90.2%。

2. 适合度和选择系数

达尔文所描述的自然选择是以"生存"来表示的，然而在自然选择的过程中最重要的是贡献给后代的相对基因数目。一定的生存能力是重要的，但是生存还不足以说明都能把基因传给后代。因此我们要通过对繁殖的估价来度量自然选择。**适合度**(fitness)就是用来度量自然选择的参数，它也被称为基因型相对的繁殖能力。适合度是一种基因型能在一定环境下生存并将其基因传给下一代的相对能力。

适合度常用 W 来表示。由于适合度是用来量度相对的繁殖能力，群体遗传学通常指定适合度为 1 时，这种基因型能产生最多的后代。其他的基因型的适合度要相对次于这种基因型。

例如基因型 G^1G^1 平均产生 8 个后代，G^1G^2 平均产生 4 个后代，G^2G^2 平均产生 2 个后代，那么 G^1G^1 基因型就具有效高的繁殖效率，因此它的适合度是 1($W_{11}=1.0$)，基因型 G^1G^2 的适合度为 4/8，W12 = 0.5，而基因型 G^1G^1 的适合度 $W_{22}=2/8=0.25$。表 25-10 说明适合度的计算。适应值可以告诉我们自然选择对一种基因型作用的大小。**选择系数**(selection coefficient)是量度对某一基因型的选择强度。也可以说是表示某一基因型在群体中不利于生存和繁殖的相对程度，常用 S 来表示，如在上述例子中选择系数 G^1G^1:S = 0; G^1G^2::S = 0.5; G^2G^2:S = 0.75.

表 25-10　三种基因型的适应值和选择系数的计算

	基因型		
	G^1G^1	G^1G^2	G^2G^2
在一个群体中交配的成体数	16	10	20
在下一代中这种基因型成体产生的后代数	128	40	40
每个交配成体所产生的平均后代数	128/16 = 8	40/10 = 4	40/20 = 2
适应值(产生后代的相对数)	8/8 = 1	4/8 = 0.5	2/8 = 0.25
选择系数	1 - 1 = 0	1 - 0.5 = 0.5	1 - 0.25 = 0.75

引自 Peter J. Russell: Genetcs, 3rd ed., Table 23.8.

3. 选择对基因频率的作用

自然选择产生许多不同的效应。有时自然选择可以排除遗传变异，有时又可以维持遗传变异。它既可以改变基因频率也可以阻止基因频率的改变。它可以产生群体中的遗传歧化，也可维持遗传的一致性。这些作用取决于基因型的相对适应性和群体中

等位基因的频率。

自然选择导致基因频率的改变可以通过表 25 - 11 的方式来计算。此方法可以用于单个座位的性状，无论这个性状是显性、隐性，共显性还是超显性。表中列了三种基因型（A^1A^1、A^1A^2、A^2A^2）以及它们最初的频率。若是随机交配的话这些基因型是适合哈迪 - 温伯格比例，开始时的频率分别为 p^2、$2pq$ 和 q^2。然后表中分别列出了三种基因型的适合度。现在假设有选择发生，而且仅某种基因型能幸存。每种基因型对下一代基因的贡献将等于开始时的基因频率乘以适合度。如 A^1A^1 对下代的贡献是 $p^2 \times W_{11}$。要注意这三种基因型的贡献相加并不等于 1。而它们的总和等于 $p^2\ W_{11} + 2pqW_{12} + q^2\ W_{22} = W$，此称为此群体的平均适应值。其次应将相对的贡献除以群体的平均适应值，使其标准化，使我们能获得选择后的相对的基因型频率。然后我们来计算选择后某种基因型所形成的基因频率（p'），此可用我们所熟悉的公式：$p' = f(A^1A^1) + \frac{1}{2}f(A^1A^2)$，最后后用 $p' - p$ 就获得由于选择而引起的基因频率的改变。一个实际计算的例子如表 25 - 12 所示。

表 25 - 11　确定由于自然选择使基因频率发生改变的方法

	基因型		
	A^1A^1	A^1A^2	A^2A^2
基因型频率	p^2	$2pq$	q^2
适合度	W_{11}	W_{12}	W_{22}
选择后频率	p^2W_{11}	$2pqW_{12}$	q^2W_{22}
选择后的相对的基因型频率	$P' = p^2W_{11}/\overline{W}$	$H' = 2pqW_{12}/\overline{W}$	$Q' = q^2W_{22}/\overline{W}$

Peter J. Russell: Genetics 3rd ed. Table 23.9

选择后基因频率 $= p' + \frac{1}{2}(H')$，$q' = 1 - p'$；选择后基因频率的改变 $= \triangle p = p' - p$；$\overline{W} = p^2\ W_{11} + 2pq\ W_{12} + q^2\ W_{22}$

表 25 - 12　当 $p = 0.6, q = 0.4$s 时由于自然选择使基因频率发生改变的计算

	基因型		
	A^1A^1	A^1A^2	A^2A^2
基因型频率	$p^2 = (0.6)^2 = 0.36$	$2pq = 2(0.6)(0.4) = 0.48$	$q^2 = (0.4)^2 = 0.16$
适合度	$W_{11} = 0$	$W_{12} = 0.4$	$W_{22} = 1$
选择后频率	$p^2\ W_{11} = (0.36)(0) = 0$	$2pq\ W_{12} = (0.48)(0.4) = 0.19$	$q^2\ W_{22} = (0.16)(1) = 0.16$
选择后的相对的基因型频率	$P' = p^2\ W_{11}/\overline{W}$ $P' = 0/0.35 = 0$	$H' = 2pq\ W_{12}/\overline{W}$ $H' = 0.19/0.35 = 0.45$	$Q' = q^2\ W_{22}/\overline{W}$ $Q' = 0.16/0.35 = 0.46$

选择后基因频率 $= p' + 1/2(H') = 0 + 1/2(0.54) + 0.27$，$q' = 1 - p' = 1 - 0.27 = 0.73$

选择后基因频率的改变 $= \triangle p = p' - p = 0.27 - 0.7 = -0.33$；$\overline{W}\ p^2\ W_{11} + 2pq\ W_{12} + q^2\ W_{22} = 0 + 0.19 + 0.16 = 0.35$

Peter J. Russell: Genetics, 3rd ed., Table 23.10

4. 选择隐性性状

很多重要的性状和大部分新的突变是隐性的且会减少适合度。当一个性状完全是隐性的，杂合子和显性的纯合子适合度为 1，隐性纯合子适合度减小时，设适合变为 1 - S，经过选择后 a 的频率（表 25 - 13）：

表 25－13　显性完全选择对隐性不利时基因频率 q 的改变

	AA	Aa	aa	合计
原来基因频率	p^2	$2pq$	q^2	1
适应值	1	1	1－S	
选择后频率	$p^2\times1=p^2$	$2pq\times1=2pq$	$q^2(1-S)=q^2-Sq^2$	$p^2+2pq+q^2-sq^2=1-sq^2$
相对频率	$p^2/(1-sq^2)$	$2pq/(1-sq^2)$	$(q^2-Sq^2)/(1-sq^2)$	

$$q'=(q^2-Sq^2)/(1-sq^2)+\frac{1}{2}[2pq/(1-sq^2)]=2pq/(1-sq^2)+pq/(1-sq^2)$$

$$=(q^2-Sq^2+pq)/(1-sq^2)$$

$$=(q^2+pq+sq^2)/(1-sq^2)$$

$$=[q(q+p)-Sq^2]/1-sq^2$$

$\because (p+q)=1 \quad \therefore q'=q-Sq^2/(1-sq^2)$

$$\Delta q=q'-q=(q-Sq^2)(/1-sq^2)-q$$

$$=(q-Sq^2)/(1-sq^2)-q(q-Sq^2)/(1-sq^2)$$

$$=[q-sq^2-q(1-sq^2)]/(1-sq^2)$$

$$=(q-sq^2-q+sq^3)/(1-sq^2)$$

$$=[-sq^2-(1-q)]/(1-sq^2)$$

一代的自然选择后，a 的频率不再是 q，而是为 q'，$q'=\dfrac{q-sq^2}{1-sq^2}$，

则 $\Delta q=\dfrac{-sq^2(1-q)}{1-sq^2}$

当 $\Delta q=0$ 时，等位基因频率不会发生进一步改变。要注意在等式中 sq^2 前有负号，因为 s，p 和 q 值总是为正值或零，所以 Δq 是负值或零。这样表明随着选择 q 值将减小。

选择也取决于群体中基因的实际频率，这是因为 Aa 和 aa 与 a 等位基因频率变异的相对比率会影响到选择降低隐性性状的效率。当隐性基因的频率相对较高时，很多纯合隐性个体存在于群体中，它们的适应值是低的，因此将导致基因频率有很大的改变。相反则基因频率改变比较小。

表 25－14 表明三个原来频率不相同的群体，群体 1 原来的频率 $q=0.9$，群体 2 的 $q=0.5$，群体 3 的 $q=0.1$；当 aa 的适合度为 0 而 AA，Aa 为 1 时，当 q 频率比较高时，如群体 1 频率改变是大的，在第一代中 q 从 0.9 降到 0.47。但 q 值小时，如群体 3，q 值改变也小，第一代从 0.1 仅降到 0.091，表明选择对隐性纯合不利时，原来的频率越大，选择的效果越明显。当 q 变得很小时 q 的改变也随之减小，因此实际上即使适合度为零，选择也很难将隐性性状从群体中消除掉。如果杂合子的适合度减小，那么基因频率的改变也比较快。因为选择时 Aa 和 aa 都起到作用。

我们还必须注意前面所说的当 q 值越大，基因频率改变也越大，这是指一般情况下，也有特殊的情况。现在我们来讨论一下下面的问题。当选择对隐性纯合不利时，一代后基因频率的改变为：$\Delta q=\dfrac{-sq^2(1-q)}{1-sq^2}$

当 q 很小时，分母近似 1，等式可写成 $\Delta q=-sq^2(1-q)$

表 25－14 比较在原来频率不同时选择对隐性表型的效率

代数	基因型（q^2aa）和等位基因（qa）频率					
	群体 1		群体 2		群体 3	
	q^2	q	q^2	q	q^2	q
0	0.810	0.900	0.250	0.500	0.010	0.100
1	0.224	0.474	0.112	0.333	0.008	0.091
2	0.103	0.321	0.063	0.250	0.007	0.083
3	0.059	0.243	0.040	0.200	0.006	0.077
4	0.038	0.159	0.028	0.167	0.005	0.071
5	0.027	0.163	0.020	0.143	0.004	0.066
10			0.007	0.08		

Peter J. Russell: Genetics, 3rd ed. 1992. Table 23.11

此式也反应 q 值的大小和 Δq 的关系，即 q 值越大，也越大，但对该式求导

$dq = (-sq^2 + sq^3)' = -2sq + 3sq^2$

令 $dq = 0$ 则 $2sq = 3sq^2$ $q = 2/3$

表明当 $q = 2/3$ 时，Δq 有极大值，也就是当 q 等于或接近 2/3 时选择的效果最明显（表 25－15）。图 25－11 说明选择对显性、隐性的作用效率也是不同的。

表 25－15 q 值不同，S 值不同时选择的效率也是不同的。$q = 2/3$ 时选择的效果最明显

Δq		S 值不同时所需的代数			
q	q'	$S = 1$	$S = 0.5$	$S = 0.1$	$S = 0.01$
0.99	0.75	1	8	38	382
0.75*	0.50	1	3	18	176
0.50	0.25	2	6	31	310
0.25	0.10	6	14	71	710
0.10	0.01	90	185	924	9,240
0.01	0.001	900	1,805	9,023	90,231
0.001	0.0001	9,000	18,005	90,023	900,230

* q 等于或接近 2/3 = 0.67 时基因频率改变最快

通过以上选择对隐性性状淘汰作用的讨论，说明基因频率改变的公式可以从在一定选择条件下基因频率发生改变的表格中得到映证。显性性状和共显性性状也同样如此。我们不再详细地推导和解释，但将有关计算公式在表 25－16 中列出，供计算时使用。

表 25－16 选择在不同情况下使基因频率的改变的计算公式

选择的类型	A^1A^1	A^1A^2	A^2A^2	基因频率的改变的计算
选择对隐性纯合不利	1	1	1－S	$\Delta q = -spq^2/(1-sq^2)$
选择对显隐性基因不利	1－S	1－S	1	$\Delta q = -spq^2/(1-s+sq^2)$
不选择显性基因	1	(1－S/2)	1－S	$\Delta q = (-spq/2)/(1-sq)$
选择对杂合有利	1－S	1	1－t	$\Delta q = pq(sp-tq)/(1-sp^2-tq^2)$
选择对杂合不利	1	1－s	1	$\Delta q = spq(q-p)/(1-2spq)$
选择是普遍性的	W_{11}	W_{12}	W_{22}	$\Delta q = pq[p(w_{12}-w_{11})+q(w_{22}w_{12})]/\overline{W}$

四、突变和选择

正如前面所讨论的自然选择减少有害的隐性等位基因频率，但当这种隐性基因的频率很低时，每代频率的改变也在减小（图 25－11）。当这种等位基因是十分罕见时，基因频率的改变十分细微。由于选择是突变压力，与等位基因频率下降相反，会不断产生新的等位基因，倾向于增加等位基因，最后达到平衡。在此过程中，通过突变，新产生的等位基因的引入可以精确地抗衡那些通过同自然选择而丢失的等位基因，从而获得平衡。尽管选择和突变持续地作用，只要不存在某些其他环境因子的影响这些等位基因的频率保持稳定。

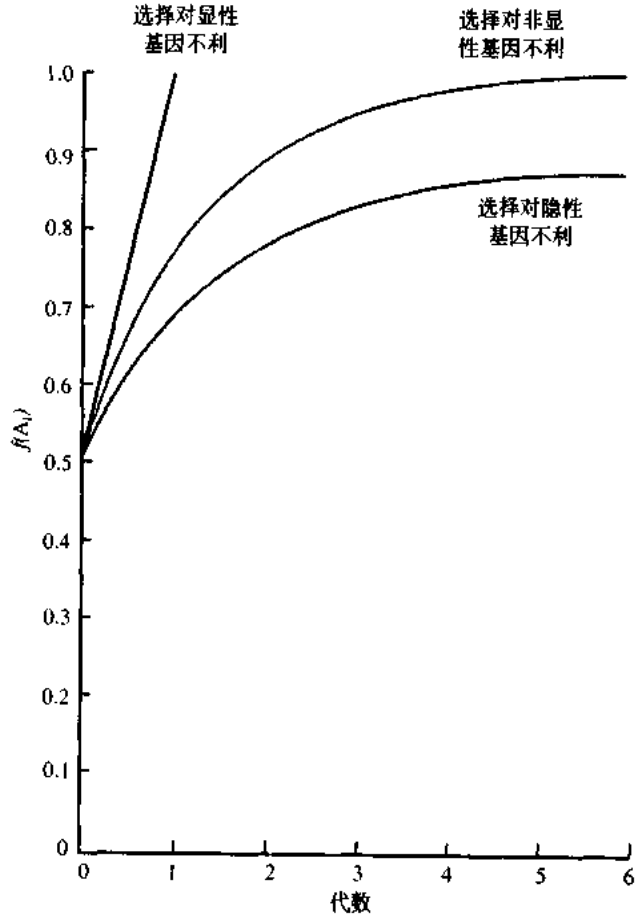


图 25－11　当选择系数相同时，不同类型等位基因频率改变的比较。A_1 是有利的基因。
（引自 Russell, P.J.: Genetics, 3rd ed., 1992, Fig23－18）

设在一个群体中选择对一个有害的隐性基因 a 不利，一代后 a 基因频率发生变化为：$\Delta q = \dfrac{-spq^2}{1-sq^2}$（表 25－15）。若 a 是十分罕见的隐性等位基因，q^2 接近于 0，分母近似等于 1。$\Delta q = -spq^2$

在此同时，若 A 突变为 a，结果使 A 等位基因增加频率为 up。因 a 的频率很低，故

其回复突变的频率可以忽略不计。在自然选择使 a 基因减少的频率和突变使 a 基因增加的频率相同时，选择和突变之间达到了平衡。

$$spq^2 = up \quad sq^2 = uq^2 = u/s \quad \hat{q} = \sqrt{u/s}$$

$\hat{q}$ 为达到平衡时 a 基因的频率。若隐性纯合体是致死的（$s = 1$），则等式变为 $q = \sqrt{u}$

例如在突变和选择达到平衡时，设隐性基因突变频率为 10^{-6}，而选择系数为 0.1，平衡时 $\hat{q} = \sqrt{10^{-6}/0.1} = 0.0032$。由于突变和选择之间达到平衡，大部分隐性有害性状以一个很低的频率保留在群体中。

对于显性基因 A 来说，平衡时 $q = u/s$，若突变率是 10^{-6}，而 s 是 0.1，达到平衡时显性基因的频率将为 $10^{-6}/0.1 = 0.00001$。可以看到在突变率和选择系数相同的情况下达到平衡时，显性基因的频率要低于隐性等位基因频率。这是由于选择对杂合状态下的隐性基因不起作用，仅对隐 性纯合体起作用。对显性基因就不同了，不论纯合还是杂合都会起作用。正是这个原因有害的显性等位基因一般都少于隐性的。

五、超显性

当杂合子比纯合子的适应性更高时也可使基因频率达到平衡。这种现象称为**超显性**（overdominance）或**杂合优势**（heterozygote adventage）。在超显性中由于杂合子的适应值高于纯合体，使得两个等位基因在群体得以维持平衡。选择将会使基因频率发生改变直到达到平衡，然后维持稳定。在这种情况下，群体中基因频率能达到平衡是依赖两种纯合体的适合度。若 AA 的选择系数是 S，而 aa 的选择系数是 t，则 $\hat{q} = f(A) = t/(s + t)$，$\hat{q} = f(A) = s/(s + t)$。

前面我们已提到人类中镰形贫血的例子，在某些群体中人的血红蛋白基因有三种基因型 *Hb-A/Hb-A*，*Hb-A/Hb-s* 和 *Hb-s/Hb-s*。在疟疾的流行区中，杂合体的适应性要优于两种纯合体。很显然这种杂合体具有抵抗这种环境的能力，或者说他们的红细胞不能维持疟原虫的生存。因此这种基因型可以抵抗疟疾，这样其适应值就高于 *Hb-A/Hb-A* 的人。有 *Hb-s/Hb-s* 基因型的人是镰形贫血的患者，选择对他们也不利。结果使疟疾的流行区存在 *Hb-s* 基因，在群体中已建立了一种平衡。现在发现 Hb-s 基因的分布区和疟疾的流行区几乎是一致的。

六、非随机交配

理想群体假设群体中的交配是随机的，但很多群体对于某些性状而言交配并不是随机的。当非随机交配发生时基因型存在的比例就难以用哈迪－温伯格定律来进行估计。

一类非随机交配是**正选型交配**（positive assortative mating）即当带有相似表型的个体优先交配。在自然群体中这种情况较为普遍。例如人类的选择性就很高。高个子女人和高个子男人婚配，矮个子女人与矮男人婚配都比较常见，比随机频率要高。**负选型交配**（negative assortative mating）是不相同表型的个体交配的概率要大于随机交配。情况和上面的例子相反。正选型交配和负选型交配的类型都不会影响群体的基因频率，但他们可能会影响基因型频率。

另外两种非随机交配的类型是**近交**（inbreeding）和**远交**（outbreeding）。在我们学习

数量遗传时近交会涉及到亲属间优先交配，近交实质是对亲属关系的正选交配。远交是非亲缘关系优先交配，实质上是负选交配。近交大部分是属于自交，此多发生在自花受粉的植物和少数自体受精的动物中，如某些种类的蜗牛。

自交会影响到基因频率的改变，可以促进等位基因的纯合（见第 7 章）。

第五节　分子进化

在近年来群体遗传学开始应用分子遗传学的技术来研究群体中的遗传变异并对某些进化的分子基础提出质疑。通过用限制性图谱和 DNA 测序方法生物学家们能在 DNA 水平列举出新的进化证据。这些研究并没有改变群体遗传的基因原理，但却提出了更为完整的详细的进化图谱。

一、DNA 序列的变异

在进化研究中一个重要的问题是在不同的基因中和在相同基因的不同部分中进化的模式和速率是怎样的？特定 DNA 序列的进化速率能通过比较由共同祖先分化出的两种不同生物的 DNA 序列来加以探讨。我们假设共同的祖先有一种单个的 DNA 序列。两种生物都由这种共同祖先演化而产生，它们的 DNA 序列经过了独立进化改变，产生了我们现在所见到的差异。例如大部分 的哺乳动物在六千五百万年以前由一个祖先进化而来。我们举一个特殊的 DNA 序列为例，如小鼠和人类的生长激素基因，我们发现二者的序列相差 20 个核苷酸。这 20 个核苷酸一定是通过了六千百万年的进化过程发生变化而歧化的。为了计算这个基因中的进化速率，我们首先要估算该基因中核苷酸的数目，也就是在这样多的核苷酸中存在着现在我发现的 20 个核苷酸的差异。通常将**进化速率**定义为每年每个核苷酸位点被另外核苷所取代的比例。为了获得变化的速率，我们就要将每个核苷酸位点核苷取代的值除以进化的年数，即两种动物分开的时间。在我们生长激素基因的例子中，改变的速率为每年每个位点取代 4×10^{-9} 个核苷酸。为数众多基因中核苷酸序列的研究揭示了基因不同部分以不同的速率在进化。回顾一下我们所讨论的分子遗传学一个典型的真核生物基因其中部分核苷酸为编码序列，将决定蛋白质中的氨基酸序列。另一部分核苷酸是非编码序列，包括内含子，前导序列和拖尾序列，这些都是只转录而不翻译的区域。5′和 3′的折叠序列也是不转录的。附加的非编码序列包括假基因，由于突变的存在，这种核苷酸序列不再产生有功能的基因产物。甚至在一个有功能基因的编码序列中也不是所有核苷酸的取代都能使蛋白质的氨基酸序列发生相应的改变。实际上许多发生密码子的第三个位置上突变对蛋白质的氨基酸序列并不起作用。因为这种突变产生的是同义密码子。

基因的不同区域所承受的进化压力不同，其进化的速度也不同，表 25－16 表明哺乳动物基因的不同部分相对的进化速率。在有功能的基因中编码序列涉及同义突变的变化速率是很高的。约 5 倍于非同义改变的变化率。同义突变并不改变蛋白质的氨基酸序列，这样高的进化速率是意料之中的事，因为这种改变不会影响到蛋白质的功能。同义和非同义突变以相同的频率产生，但在编码序列产生的错义突变常因有损适应而被自然选择所淘汰。相反同义突变不造成损害，因此它们得以保留。

注意一下表 25－17 在有功能基因的 3′端侧翼区的进化速率也很高。与同义突变相似，3′端侧翼区的序列尚不知它对氨基酸序列有何作用。通常是对基因表达有影响；

表 25－17 哺乳动物基因的 DNA 序列中不同部分的相对进化速率

序列	进化速率($\times 10^{-9}$)
有功能的基因	2.32
5′侧翼区	2.32
前导区	1.74
同义编码序列	4.65
非同义编码序列	0.88
内含子	3.70
拖尾序列	1.88
3′侧翼区	4.46
假基因	4.85

Peter J. Russell: Genetics, 3rd ed. 1992, Table 23.14

发生在此区的大部分突变将不会被自然选择所淘汰。内含子的进化率也很高，情况和同义突变及 3′侧翼区相似。虽然内含子的序列一般并不编码蛋白，但它必须经过剪接才能形成成熟的 mRNA，进而翻译成有功能蛋白。内含子中的少数序列负责剪接，包括内含子 5′和 3′端的保守顺序以及分枝点序列(见第 14 章)结果是内含子中的变化不是所有都被保留，因此其进化速率比同义突变及 3′侧翼区要略低一些，但比编码区仍然高得多。

5′侧翼区的进化速率并不高，虽然此区既不转录，也不翻译，但它含有基因的启动子，因此它对于基因的表达是十分重要的。重要的保守序列如 TATA 框就位于此处。保守序列的突变可能会妨碍基因转录，这样会损害生物的适应性；自然选择会淘汰这些突变而使该区的进化率保持在较低的水平上。

前导区和拖尾区的进化速率都比 5′侧翼区的还要低。虽然这两个区域都不翻译，但要转录，并为 mRNA 的加工以及 mRNA 附着到核糖体上提供信号。在这两个区的核苷取代是很有限的。编码序列中的非同义突变的进化速率最低。这些核苷酸发生突变会改变蛋白质中氨基酸的序列，这种突变大部分被自然选择所淘汰。

表 25－17 最后一栏中没有功能的假基因进化速率最高。例如人类的珠蛋白假基因，其核苷酸的变化率是有功能珠蛋白基因编码序列的 10 倍。这是由于这些假基因不再编码蛋白；由于这些基因的改变并不影响人的适应性，因此也不会被自然选择所淘汰。如果我们发现了某段序列有很高的进化率，那么意味着此序列可能没有什么功能。

在研究核苷酸序列中一个很有趣的发现是，虽然有些同义突变核苷酸的进化率要比非同义突变的要高得多，但比假基因的还是要低一点。这表明同义突变的适应性也并不完全相同，自然选择可能对它们的偏爱也不同。这一假说通过发现编码序列中同义密码子的使用率不同而得到加强。如亮氨酸有 6 个不同的密码子(UUA，UUG，CUU，CUC，CUA 和 CUG)，但在细菌中 60% 的 Lue 的密码子是用 CUG，在酵母中有 80% 是用 UUG。由于同义密码子对于特定氨基酸来说都是一样的，那么必然是选择对某些密码子有所偏爱。某些同义密码子与不同的 tRNA 的配对也是如此。因此同义突变虽然并没有改变氨基酸，但可能改变了 RNA。对不同的 tRNA 进行研究表明**同工受体 tRNA**(isoacceptor tRNAs)(指不同的 tRNA 可以接受相同的氨基酸)的量不同，而最丰富的 tRNAs 是和最常用的密码子配对，选择的偏爱可能是由于某些密码子相应的 tRNA 更为丰富。密码子与反密码子的结合能量可能也是不同的，这是由于涉及到不同碱基的

配对，此可能会受到自然选择的作用。

另外不同功能的基因进化速率也不同（表 25－18）例如人类的催乳素基因中非同义核苷酸的取代率要比哺乳动物组蛋白 H4 基因的高 300 倍。

表 25－18　不同的哺乳动物基因的 DNA 序列中相对进化速率

基因	非同义突变进化速率（$\times 10^{-9}$）	同义突变进化速率（$\times 10^{-9}$）
组蛋白 H4	0.004	1.43
胰岛素	0.16	5.41
催乳素	1.29	5.59
α－珠蛋白	0.56	3.94
β－珠蛋白	0.87	2.96
白蛋白	0.92	6.72
甲胎蛋白	1.21	4.90

Li, W. *et al.*: Mol. Bio. Evol. 1958, 2: 150－174

二、DNA 长度的多态性

除了核苷酸序列通过核苷酸取代而进化以外，基因中核苷酸的数目也常发生变异。这种变异称为 **DNA 长度的多态性**（DNA length polymorphisms），它是通过缺失或增加一段相对短的核苷酸序列而产生的。例如在黑腹果蝇的乙醇脱氢酶基因中发现了 DNA 长度的多态性。Martin Krietman 测定了这个基因的 11 个拷贝，他发现除了核苷酸序列存在变异外，这 11 个拷贝中有 6 个插入和缺失。所有这些都有在内含子和 DNA 的侧翼区内，在外显子中未发现有这种情况。外显子中的插入和缺失常会改变读框，因此它们将受到选择的作用。结果插入和缺失通常只发生在 DNA 的非编码区。然而某些插入和缺失在有些基因的编码区已有发现。另一种 DNA 长度多态性涉及到特殊基因拷贝数多少的差异。例如在个别果蝇中核糖体基因拷贝数发生广泛的变化。棒眼基因的拷贝数在突变体中也常不同。转座子的拷贝数在个体中变化也很大，可能引起某些 DNA 长度的多态性。

三、多基因家族的进化

在真核生物中我们常会发现基因的多个拷贝，这些拷贝的顺序都相同或相似。这样的一组基因为多基因，称为**多基因家族**（multigene family）。这样的一组基因是由一个祖先基因通过重复进化而来，基因家族的成员可以彼此形成基因簇或分居于不同的染色体上。

例如珠蛋白的基因家族就是一个多基因家族，它们由编码珠蛋白分子多肽链的基因组成。在人类的第 16 号染色体上发现了 7 个 α－珠蛋白基因，在第 11 号染色体上发现了 6 个 β－珠蛋白基因。在动物中也发现了珠蛋白基因。甚至在植物中也发现了珠蛋白样基因，表明这是一个非常古老的基因家族。在多种动物中几乎所有有功能的珠蛋白基因结构都相同，由 3 个外显子组成，中间间隔着两个内含子。但珠蛋白基因的数量和次序在各种动物中是不同的。图 10－31，32 表示几种动物的 β－珠蛋白基因家族。由于所有的珠蛋白基因的结构和顺序都是相似的，因此它们存在着一个原始的珠蛋白基因（多半和现在存在的肌红蛋白基因相关）。经重复和歧化而产生了原始的

α－珠蛋白基因和β－珠蛋白基因。植物的豆血红蛋白基因是和珠蛋白基因相关的。植物豆血红蛋白基因存在着很多原始的类型，它有一个额外的内含子，尚不清楚是由一个额外的内含子插入到植物的相应基因中，还是其他的进化路线中丢失此内含子。

通过哺乳动物肌红蛋白单个基因为我们提供了追踪珠蛋白基因的线索。它是约在8亿年以前和珠蛋白进化路线上分开的。肌红蛋白基因的组成和珠蛋白基因相似。因此我们可以将三个外显子结构看成是它们共同的祖先。

某些原始的鱼类只有单个类型的珠蛋白链（图25－12），因此它们必然是在珠蛋白基因尚未发生重复前就歧化了出来。这个基因重复后经突变形成α－和β－两种不同的珠蛋白基因，在某些两栖动物中就含有α和β连锁的珠蛋白基因。即幼体型和成体型。后来进一步重复，在和哺乳动物中形成了α－珠蛋白家族和β－珠蛋白家族。重复在进化中是常发生的。在某些人类群体中珠蛋白基因的拷贝数是有变化的。例如大部分人在16号染色体上有2个α－基因，但有些个体在此染色体上只有1个α－珠蛋白基因。而另一些个体有的甚至有3个、4个α－珠蛋白基因。此表明在多基因家族中基因的重复和缺失是恒定的进行过程。基因重复和缺失常常是由于在交换过程中顺序错排所致，这种交换的过程称为**不等交换**。重复也可以通过转座而产生。随着基因的重复，基因拷贝的分离可能经受序列的改变。在有的情况下，突变会使基因拷贝变得无功能，从而产生假基因。在另一些情况下，核苷酸序列的改变也可导致基因产生的蛋白具

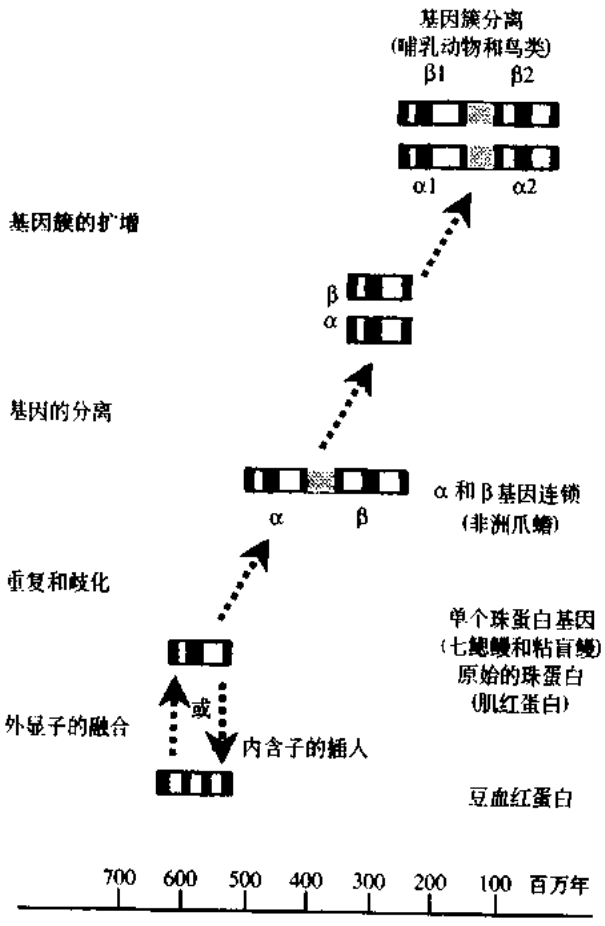


图25－12 所有珠蛋白的基因通过一系列的重复，转座和突变而发生进化。（仿B.Lewin:Genes VI.1997，Fig.23－7）

有不同的功能。

另一个有趣的例子是乳酸脱氢酶同工酶的进化。人们用电泳，比较 DNA 序列等方法研究了 LDH 不同亚基在不同生物中的分布及表达，描绘出了一种分子进化的途径。

人们发现在最原始的脊椎动物——无颌类中（七鳃鳗）中仅有 LDH-A 一种亚基存在，而在软骨鱼中已发现有两种亚基 LDH-A 和 LDH-B 存在。进化到硬骨鱼类开始出现了第 3 个亚基 LDH-C，不仅鱼类以下的动物无 C 亚基，且软骨鱼及原始的硬骨鱼肺鱼类都没有出现 C 亚基，直到鲟形目才开始出现 C 亚基。在低等的硬骨鱼中 *ldh-C* 基因在各种组织中都广泛表达。随着进化地位的升高，有的仅在肝中表达，另一类仅在眼中表达，显示了基因表达调控的组织特异性；在两栖类的黑爪异鲵中（*Xenobius melanonychus*）发现仅在雄性的肝脏中表达，不仅显示了组织的特异性还显示了性别的特异性，可能受到雄性激素的调控。进化到鸟类和哺乳类 *ldh-C* 仅在精子中表达，这不仅具有组织和性别的特异性，而且显示了发育阶段的特异性。以上表明基因的重复和特化是新基因产生的必要途径。通过基因克隆和分子杂交及测序，已证明鸽子的 *ldh-C* 基因和 *ldh－B* 基因连锁，且同源性较强；人和小鼠的 *ldh-C* 和 *ldh-A* 连锁，且同源性大于 B.C 之间。表明哺乳动物和鸟类的 ldh－C 基因来源虽不相同，但都是由于相关基因二次重复和突变所形成。LDH 的进化反映了在进化过程中不仅有编码区的改变，而且调控区也在发生改变。LDH-C 逐步具有组织的特异性，性别的特异性和发育阶段的特异性（图 25－13）。

四、线粒体 DNA 的进化

前面我们已介绍了线粒体基因组（见第 20 章）。动物的 mtDNA 约 15kb，它们编码 2 个 rRNA，22～23 个 tRNA 和 10～12 个蛋白。近年来很多的研究者都发现 mtRNA 中的序列有变化。他们推测 mtRNA 的进化和核 DNA 的进化是不同的。例如动物 mtDNA 中的核苷酸顺序的进化速率要比核基因的编码顺序快得多。其实哺乳动物的 mtDNA 进化速率要比其核 DNA 快 5～10 倍。但从动物 mtDNA 所获得的进化速率并不代表所有真核生物的。如不知何故植物 mtDNA 的进化速率就要比核 DNA 进化速率慢得多。

还不清楚为什么动物的 mtDNA 进化速率如此之快，有可能和 mtDNA 顺序的高突变率有关。在 mtDNA 存在较高突变率是因为其缺乏修复机制或者由于在 mtDNA 复制中的 DNA 聚合酶存在较多误差倾向。动物 mtDNA 中有较高进化速率的另一种可能是选择压力在核基因中淘汰了很多突变，而在线粒体中这种压力松弛了。再者由 mtDNA 编码的蛋白质，tRNAs 和 rRNAs 的改变可能比核基因编码的蛋白，tRNAs 和 rRNAs 的改变对个体适应性的损伤更少一些。

无论什么原因加速 mtDNA 进化，这些序列迅速改变的事实使得人们想到应用 mtDNA 来分析亲缘关系很近的生物。例如将 mtDNA 顺序分析用于研究人类、黑猩猩、大猩猩、猩猩及长臂猿之间精确的进化关系。

mtDNA 和核 DNA 不同，所有 mtDNA 都是直接由母体遗传的，它位于细胞质里，而且只通过卵细胞带到合子中。mtDNA 没有经历过减数分裂，而且所有后代中的 mtDNA 的顺序应是相同的。这种遗传方式形成了母系传递，这样在一个群体中一个家族的 mtDNA 的顺序是相同的。

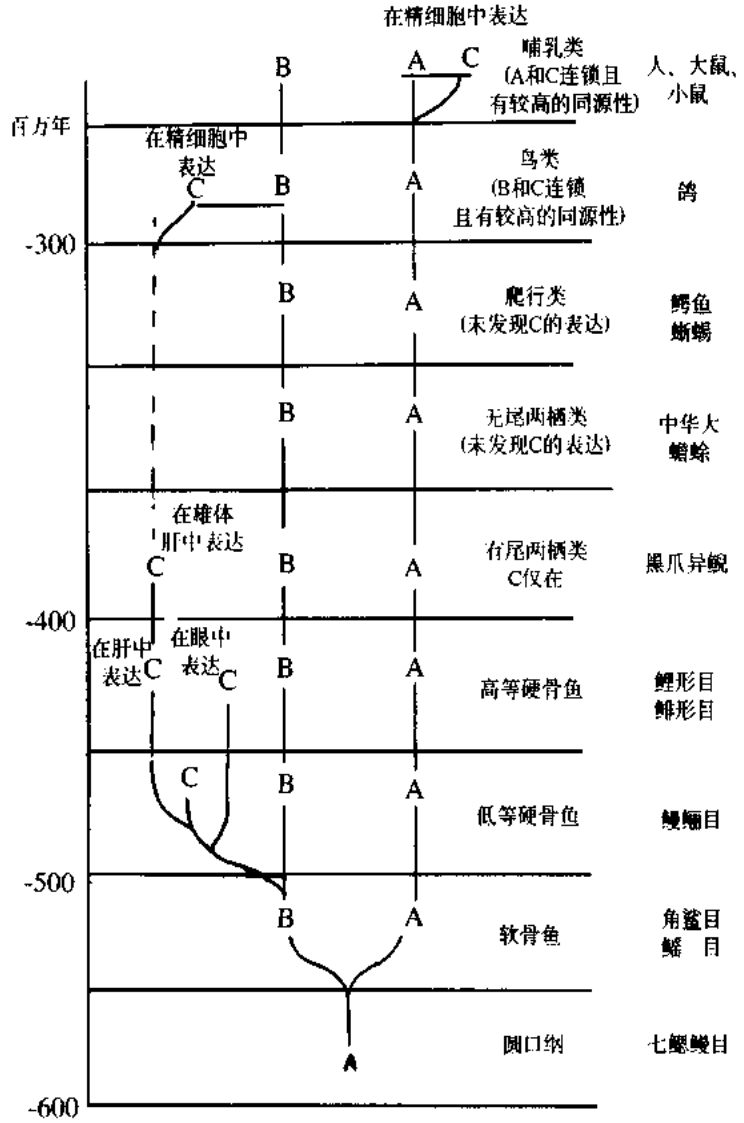


图 25-13 LDH 基因在脊椎动物中的进化

五、协同演化

在 DNA 序列变异研究中人们惊奇地发现某些分子的力量维持着基因多拷贝中的一致性序列。这种现象为**协同进化**(concerted evolution)或**分子驱动**(molecular drive)。前面我们已经讨论过在真核生物中某些基因存在着多拷贝。例如在复杂的生物中核糖体 RNA 基因存在几百或几千个拷贝。这些多拷贝是由于重复而产生。随着基因的重复，单个拷贝已可能发生突变和歧化。选择能限制编码区的突变。但如果有很多拷贝存在的话，预期也可能有歧化的发生，特别是在非编码序列中。但实际的情况和预计的相反，许多研究结果表明，在基因不同拷贝中的核苷酸序列常常是完全同源的。此外非编码序列也是相同的。表明单纯的选择是不起作用的。在亲缘关系很近的物种中存在着相同基因，这些基因的序列也是同源的，但常在某些物种中发现同源序列的差异。

这些观察产生了这样的结论，某些分子的力量持续地维持着物种中相同序列的多

拷贝。在相同的时间内，可迅速地形成物种间的差异。协同进化机制现在还不十分清楚，但它对于基因的演化是十分重要的。

六、RNA 和 DNA 序列显示了进化的相关性

在短短的几年中分子遗传学就为破译生命的进化史提供了有力的工具。由于进化的本质是遗传的改变，遗传的相关性对于进化树的构成来说是十分重要的。然而直到最近，还不能直接地审查基因。过去进化生物学家完全相信通过表形的比较来推测遗传的相似性和差异。他们假定若表型相似，控制这个表型的基因也相似，若表型不同，则相关的基因也不同。这样表型就被用于进化的研究。原先一些所谓的"表型"常常是由大量的粗糙解剖特征所构成。后来对行为，超微结构和生化特征也进行了研究。这些特点被有效地用于很多动物、植物种群进化树的构建，为现代大部分进化的研究打下了基础。

然而这种性状的研究是有限的。有时在生物中表型虽相似，而亲缘关系却很远。例如天真的生物学家试图构建一生物进化树，其依据是翅膀的有无，那么就会把鸟类、蝙蝠和昆早放在一个进化树中，因为它们都有翅膀。情况十分明显这种生物亲缘关系是很远的。它们除了翅膀以外很多方面都不相同，即使是翅膀本身也不同。当然这是个极端的例子，说明有时表型会使我们在进化关系上被引入歧途，表型的相似性并不能完全反映基因型的相似性。

另一个和表型有关的问题是：不是所有的生物都具有很多容易研究的表型性状。如研究细菌间的进化关系总是有很多问题，这是因为细菌很少有与亲缘关系远近有关的性状。第三个问题是有些生物之间很少有共同的性状如动物、植物和微生物之间，我们怎样来比较呢？

现在我们来看看 DNA 测序和限制性片段长度的多态性如何为我们提供有关 DNA 序列的信息。DNA 序列为我们提供多而精确的有关进化相关性的信息。它们可以允许人们对不同生物进行直接比较。也易于定量。而所有的生物也都具有 DNA 或 RNA，至少各种生物都具有一些共同的基因，如 tRNA 基因，rRNA 基因，及少数蛋白质基因。由于这些歧化很重要，所以很多生化生物学家转向研究和进化关系相关的 DNA 序列，并构建进化树。

有一种情况是 DNA 的资料为我们提供的进化关系的新信息，使我们了解了生命的初步的歧化。在很多年以前生物学家把所有的生物分为两大类，即植物和动物。当发现了更多的生物时，对生物的性状考察得更为细微，这种简单的二分法就不能适应了。后来认识到按照细胞的结构可将生物分成为原核生物和真核生物两类。近代 Whittaker 将生物分为 5 界：原核生物、原生生物、植物、真菌和动物。

在过去几年中 RNA 和 DNA 的顺序被用于几乎所有生物进化路线的研究。其中的工作之一是 Norman Pace 及其同事以 16S rRNA 的顺序（各种生物都具有）为基础构建了一个进化树（图 25－14）。这个进化树反映了三个主要的进化群：真细菌（传统的原核生物），真核生物和古细菌（知道得较少的一群细菌）。真细菌和古细菌二者都是原核的，但二者之间与真细菌和真核生物之间一样也存在差别。真细菌和细菌在表型上进化的差别不大明显。这在比较它们的核苷酸序列后这种差别就变得很清楚。其他的基因的序列，包括 5S rRNA，大的 rRNA 及某些基本蛋白的基因支持在生物间存在这三大进化群。

DNA 顺序也被用于人类进化的研究。在人类群体中，身体的大小，形态，面部

特征，肤色等遗传差异相对较小。例如分析 mtDNA 顺序表明在两个群体的差异约为 3.3‰。其他灵长类动物存在很大的差异。例如两个巨猿亚种之间的差异大到百分之五。此表明人类各种群之间的亲缘是很近的。然而在人类不同种群中也存在某些遗传的差异。在居于各大洲的人群之间没有发现很大的差异。但这种大的差异在非州人群中是存在的。很多专家解释这些发现意味着人类在非洲经历了起源和早期的进化。在非洲群体产生很多差异之后得到了发展，少数人离开了非洲，到达世界各地，产生了各种其他人群。这个假说叫做"**非洲起源说**"。 mtDNA 和核基因的序列资料是这一学说的证据。虽然这一学说并未被广泛接受，但 DNA 序列的资料在研究人类的进化和其他种群的进化中无疑有重要的作用。

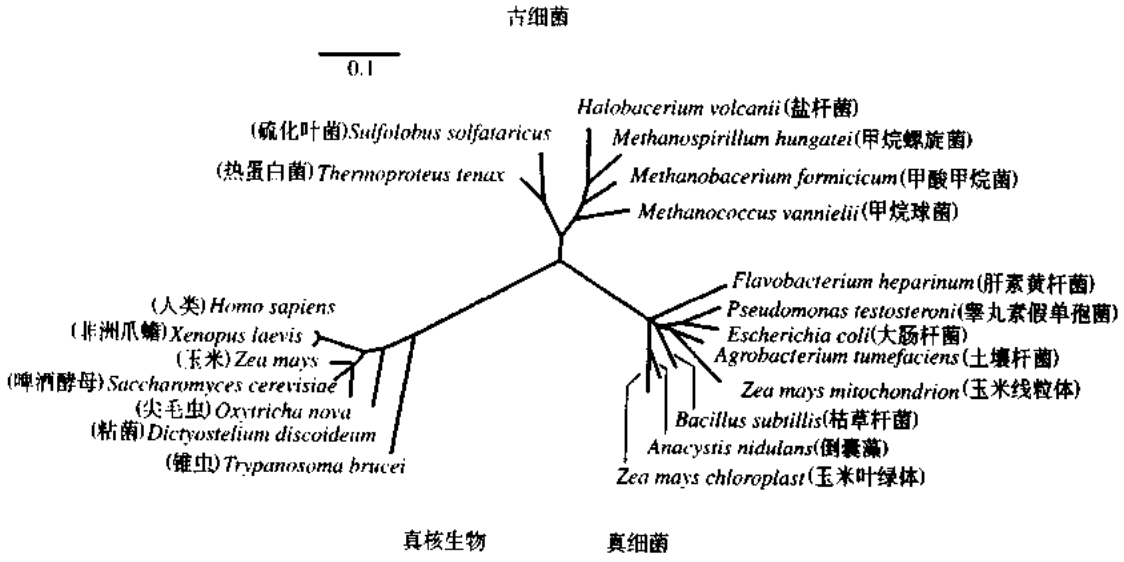


图 25 - 14　通过 16S rRNA 序列的比较所绘制的三界生物进化树。(引自 Pace, *et al*.: Cell, 1986, 45: 325 - 326)

习　　题

（附参考答案）

第一章习题

1.1 在番茄的果实中，红色对黄色是显性的，若把纯合红色番茄与纯合黄色番茄杂交，则下列情况中，番茄为何颜色？

(1) F_1 代；　(2)F_2 代；　(3)F_1 与红色亲本杂交的后代；　(4)F_1 与黄色亲本杂交的后代。

1.2 红色番茄与黄色番茄杂交，其后代中一半是红色，一半是黄色，判断亲本的基因型。

1.3 玉米中，显性基因 *A* 决定种子有颜色而 *a* 决定种子无颜色，隐性基因 *wx* 产生蜡质淀粉(waxy starch)而 *wx* 产生正常淀粉。两个基因不连锁。当 AaWxWx 的植株测交时，写出后代的基因型及表型。

1.4 F_2 代植株中有色与无色的比为 3∶1，若随机挑有色植株并将其自交，则其后代中出现多种类型植株的机率是多少？

1.5 豚鼠中粗糙皮毛的基因 *R* 对光滑的基因 *r* 是显性的。粗糙型豚鼠与光滑型杂交，产生 8 只粗糙型，7 只光滑型 F_1 后代。(1)试写出亲本和后代的基因型。(2)若一粗糙型 F_1 小鼠与粗糙型亲本杂交，后代表型如何？

1.6 在牛中，无角(*P*)对于有角(*p*)是显性的。一只公牛当与有角母牛杂交时可产生有角后代；当与无角母牛 B 杂交可产生有角后代；当与有角母牛 C 杂交可产生无角后代。(1)判断四只牛的基因型。(2)这三次杂交中其后代的分离比理论上各为多少？

1.7 某种植物中，紫色花对白色花是显性的。当某紫色花植株自花授粉时，产生 28 个紫花各 10 个白花后代，问：后代中紫色花是纯种的比例是多少？

1.8 两只黑色母鼠与同一棕色雄鼠杂交。母鼠 X 产生 9 只黑鼠，7 只棕鼠，母鼠 Y 产生 14 只黑鼠。试分析决定黑，棕毛色的遗传机制及三只鼠的基因型。

1.9 当被病毒侵染时，豆类植株会有不同症状。有的仅是局部损伤，有的为全部损害。现有一遗传分析如下：一局部损伤植株与一全部损伤植株杂交，F_1 代中只有局部损伤植株生成，F_2 代中有 785 个局部损伤，269 个全部损伤。问：(1)豆类植株此种不同的遗传基础是什么？(2)把上述与实验有关的基因命名，并设计一种测交实验证明你的假设。

1.10 小鼠的毛由等位基因 *B*、*b* 决定。其中 *BB* 和 *Bb* 决定产生黑色，*bb* 产生白色。一农民买了一对黑色小鼠，其杂交产生 6 黑 2 白的子代。他卖出白色小鼠，为产生更多的白色小鼠，他向你咨询：(1)若在 F_1 代中随机杂交，F_2 代中白色的机率？(2)若用 F_1 雄性个体与亲本雌性个体杂交，后代中产生白色个体的机率？(3)产生白色个体最多的方法是什么？

1.11 豌豆中，紫花 *P* 对白花 *P* 是显性的皱状豆荚 *S* 对光滑型 *s* 显性。现有一白色皱状的纯合植株与一紫色光滑型纯合植株杂交，判断下列各情况的表型：(1)F_1 代；(2)F_2 代；(3)F_1 个体与白色皱型亲本杂交产生的后代；(4)F_1 个体与紫色光滑型亲本杂交产生的后代。

1.12 试用析法算出下列杂交产生后代的比例(*p*、*s* 意义同上题)(1) *PPss* × *ppSS*；(2) *PpSS* × *ppss*；(3) *PpSs* × *PpSS*；(4) *PpSs* × *PpSs*；(5)*PpSs* × *Ppss*；(6) *PpSs* × *ppss*。

1.13 在南瓜中白色果(*W*)显性于黄色果(*w*)，扁形果(*D*)显性于圆形果(*d*)。试判断下列各情况中亲本的基因型。(1)白色扁形 × 黄色圆形产生 1/2 白扁，1/2 白圆后代；(2)白色圆形 × 白色圆形，产生 3/4 白圆，1/4 黄圆后代；(3)黄色扁形 × 白色圆形，只产生白扁后代；(4)白色扁形 × 黄色圆形，产生 1/4 白扁，1/4 白圆，1/4 黄扁，1/4 黄圆；(5)白公扁形 × 白色圆形，产生 3/8 白扁，3/8 白圆，1/8 黄圆，1/8 黄圆。

1.14 基因 *a*、*b*、*c*、独立分布，它们相应其等位基因 *A*、*B*、*C* 是隐性分布的，现有两个 *AaBbCc* 个体杂交。求：(1)后代中表现出三种显性性状的个体的比例；(2)后代中表现出三种显性性状的纯合体的比例？

1.15 豌豆中高茎(*T*)显性于矮茎(*t*)，绿豆荚(*G*)显性于黄豆荚(*g*)，光滑种皮(*S*)显性于皱缩种皮(*s*)，现有一矮茎，绿豆荚，皱缩种皮的纯合体与一高茎，黄豆荚光滑种皮的纯合体的杂交：(1)F_1 的表型？(2)F_2 的表型？(3)F_1 与矮，绿黄，皱亲本回交产生后代的表型？(4)F_1 与高，黄，平亲本回交产生后代的表型？

1.16 两纯合小麦杂交，它们有六对基因互不相同，这些基因独立分布并产生独立的表型效应，F_1 自交产生 F_2 代。(1)F2 可能有多少种基因型？(2)F_2 中六个基因座位上都纯合的个体的比例？

1.17 鼠的毛色由几对基因控制，显性基因 A 产生黄色条纹，纯合子不产生条纹，显性基因 B 产生黑色，隐性基因 b 产生棕色。基因型 C 使斑点遍布全身，纯合体仅在耳尖，吻端，四肢末端，尾端出现黑色称喜玛拉雅白化。(1)一纯合黑色鼠与一纯合的棕色，有黄色条纹，的喜玛拉雅鼠杂交，F_1，F_2 的表型如何？(2)黑色且有黄色条纹的 F_2 个体中其基因型是 *AaBBC* 的比例是多少？(3)喜玛拉雅鼠的 F_2 个体中显棕色的比例是多少？(4)有黄色条纹的 F_2 的个体中显黑色的个体的比例？

1.18　某种狗中毛色纯色显性于杂色，一纯合的纯色狗与一纯合杂色狗杂交：(1)产生小狗中是杂色的比例？(2)若生出四只小狗，它们都是纯色的比例？

第二章习题

2.1　间期相应于细胞周期的_____部分。

(1)有丝分裂；(2)S；(3)G1 + S + G2；(4)G1 + S + G2 + M。

2.2　通过着丝粒连结的染色单体叫_____。

(1)姐妹染色单体；　(2)同源染色体；　(3)等位基因；　(4)双价染色体。

2.3　有丝分裂和减数分裂在_____方面有所不同。

(1)染色单体　(2)源染色体　(3)双价体　(4)着丝粒　(5)纺锤体

2.4　判断并解释下列句子的正误：(1)在任何生物中的体细胞中，染色体的形态总是相似的。在有丝分裂中，染色体的分离与姐妹染色单体分离均发生在后期，并分处于两个核。每个核的染色体与亲代一样多。(2)在合子中，任何一个染色体均可与同一细胞的其他染色体联会。

2.5　回答下列问题并解释(1)减数分裂能发生在单倍体物种中吗？(2)减数分裂能发生在单倍体个体中吗？

2.6　真核生物的一般过程是(1) 1N - 减数分裂 - 2N - 受精 - 1N；(2) 2N - 减数分裂 - 1N - 受精 - 2N；(3) 1N - 有丝分裂 - 2N - 受精 - 1N；(4) 2N - 有丝分裂 - 1N - 受精 - 2N。

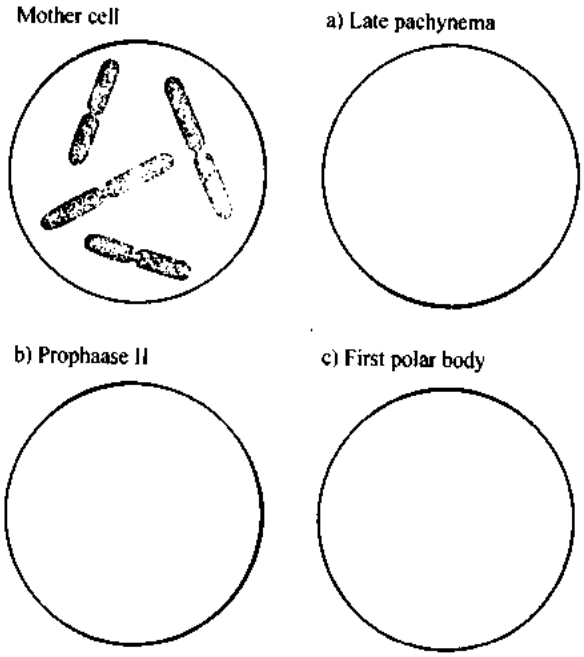
2.7　下列叙述中哪个是正确的？(1)配子是 2N，合子是 1N。(2)配子，合子都是 2N。(3)配子细胞和体细胞中的染色体数目相同。(4)合子细胞和体细胞中的染色体数目不同。(5)单倍体生物有单倍体合子。

2.8　下列各项哪项不发生在减数分裂的前期Ⅰ：(1)染色体浓缩(2)同源染色体配对(3)交叉形成(4)端化(5)同源染色体的分离。

2.9　当发生下列事件时，应分别处于有丝分裂，减数分裂的何期？(1)染色体处于纺锤体中央的平面上；(2)染色体从赤道板向两极移动。

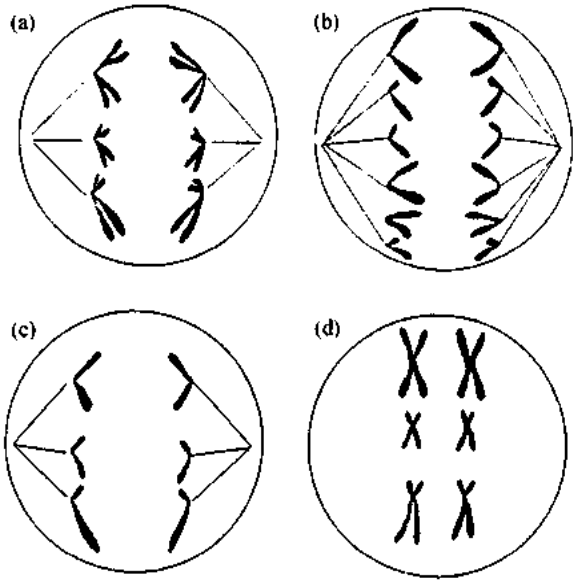
2.10　一个两倍体的将进行减数分裂的卵母细胞如图所示，画出下列各情况下的染色体。

(1)粗线期的最后阶段　(2)前期Ⅱ　(3)一极体



习题 2.10 图

2.11　下图所示的细胞均来自同一个体(哺乳动物)，判断所发生的细胞分裂事件，并解释。另外判断此个体的性别及其二倍体中所含的染色体数目。



习题 2.11 图

2.12 在遗传学上，有丝分裂，减数分裂哪一个更有意义？为什么？

2.13 某二倍体生物有三对染色体，假定其从母亲处获得染色体 A.B.C，从父亲外获得 a，b，c，回答下列问题（假定无交叉产生）：(1)此个体的配子中的染色体全来自母亲的比率？(2)此个体的配子中的染色体既来自母亲又来自父亲的比率？

2.14 在有丝分裂中期，一模式哺乳动物的正常二倍体细胞通过细胞学方法检查其染色体组。发现有一条短的染色体，两条中等长度的染色体，3 条长的染色体。试解释怎么会有这样一套染色体？

2.15 判断下面句子的正误并解释(1)在快速分裂的表皮细胞中，通过合适的染色方法可看到其减数分裂期的染色体。(2)来自同一个男子的所有精子在遗传上是相同的。

2.16 马是二倍体，共有 64 条染色体。驴也是二倍体，共有 62 条染色体。骡是由母驴和雄马杂交而生成的，它能生存但无生育能力，骡的细胞中有多少条染色体？

第三章习题

3.1 某动物体细胞中有 5 对染色体，其染色体 1，2，3，4，5 来自父亲，1′，2′，3′，4′，5′来自母亲，则其生殖细胞中既含有父亲又含有母亲染色体的比例是多少？

3.2 在男人中，每条性染色体来自其祖父一代中的哪个个体，(在下列空格处填 √或×)

	外祖母	外祖父	祖母	祖父
X 染色体				
Y 染色体				

3.3 在果蝇中，白眼是性连锁的性状，突变的白眼等位 w 隐性于野生型的红眼基因 w^+

(1)若白眼雌蝇与红眼雄蝇杂交，F_1 代的雌蝇与其父本回交，F_1 代的雄蝇与母本回交，这两次杂交后代的眼色会怎样？(2)白眼雌蝇与红眼雄蝇杂交，源于此杂交的 F_2 代自交产生 F_3 代，其眼色如何？

3.4 一种人类色盲是由性连锁的隐性突变基因 c 造成的。一正常女子(其父是色盲)与另一正常男子(其父是色盲)结婚，其后代是色盲的比例是多少？

3.5 人类红绿色盲是 X 上连锁的隐性基因造成的，一正常女子与一色盲男子生出一个色盲女子，此女子的基因型如何？

3.6 人的红绿色盲，由 X 连锁的隐性基因 c 造成的，而白化病是常染色体上一隐性基因造成，现有一对纯合体夫妇，女子无色盲但有白化病，男子是色盲但不是白化病，则其后代的表型如何？

3.7 果蝇中短翅(vg)隐性于正常的长翅(vg^+)。此对基因位于常染色体上。决定白眼的基因位于 X 染色体。现有一纯合白眼长翅的雌果蝇与一纯合白眼短翅雄蝇杂交，求

(1)F_1 代表型及相应比例　　(2)F_2 代的表型及相应比例

3.8　两红眼长翅果蝇杂交，产生后代如表所示，则此两果蝇的基因型如何？

	雌	雄
红眼长翅	3/4	3/8
红眼短翅	1/4	1/8
白眼短翅		3/8
白眼短翅		3/8

3.9　家禽中，显性的性连锁基因 *B* 产生带状羽毛，而其等位基因 *b* 纯合产生非带状羽毛。现有一非带状羽毛的公鸡与一带状羽毛的母鸡杂交。

(1)F_1 代情况如何？(2)若 F_1 代母鸡与其父本杂交，其后代情况如何？(3)若 F_1 代公鸡与其母本杂交，其后代情况如何？

3.10　某男 A 因缺少珐琅质而牙齿为棕色，A 与一正常女子结婚，其女儿均为棕色牙齿而其儿子都正常。A 的儿子一正常女子结婚，其子女都正常。A 的女儿与正常男子结婚，其后代有一半是棕色牙齿，试解释。

3.11　人类对苯硫脲味觉的区别决定于常染色体上一对等位基因，无味觉的性状隐性于有味觉，一对均有此味觉的夫妇生下一个无此味觉的孩子，他们下一个孩子的味觉会怎样？

3.12　囊性纤维化由常染色体上一隐性基因决定，两个没有囊性纤维化的父母生出两个囊性纤维化和三个没有纤维化的孩子，问：(1)他们下一个孩子是纤维化的可能性为多少？(2)他们认为没有纤维化的孩子是杂合体，你认为没有纤维化的孩子是杂合体的可能性为多少？

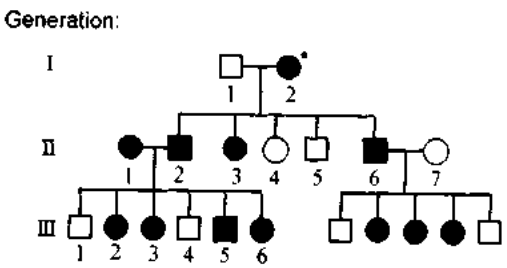
3.13　基因 *a* 位于 *X* 染色体，基因 *B*, *C*, *D* 分别位于三个常染色体。一 *AABBCCDD* 女子与一个 *aYbbccdd* 男子结婚：(1)在 F_1 中得到 *A* 的几率？(2)在 F_1 得到 *a* 性状男子的几率？(3)在 F_1 中得到 *A* - *B* - *C* - *D* - 女子的几率？(4) F_2 代将出现多少种基因型？(5) F_2 代中四对基因均为杂合的比率？(6) F_2 代中得到下列基因型的比率：①*A* - *bbCCdd*(女)；②*aYBBCcDd*；③*AybbCCdd*；④*aabbCcDd*(女)。

3.14　果蝇白眼基因 *w* 是一隐性性连锁。现从一纯合白眼雌蝇中分离到一个在第一次减数分裂时其性染色体没有分离的卵细胞。将此细胞与另一红眼雄蝇的精子受精，产生的后代分别与父本，母本回交，则回交产生的基因型，表型如何？(最初母本基因型为 *ww*，父本为 W^+Y)

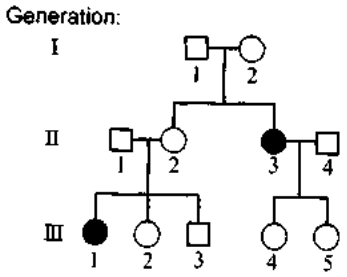
3.15　果蝇中截毛基因(bobbed gene) bb^+ 位于 X 染色体，但不同于大多数的 X 连锁基因的是，Y 染色体上携带此基因，突变体基因 *bb* 隐性于 b^+b^+。一个雌性截毛果蝇在其第一次减数分裂时其性染色体分离。它的卵细胞与一野生型雄蝇的精子受精，产生野生型 F_1。F_1 中的雌蝇与一截毛雄果蝇杂交，其后代的基因型和经例如何？

3.16　在谱系中，罗马数字表示代数，阿拉伯数字表同一代中的个体，根据下图回答下列问题(假定是单基因遗传)(1)首先排除 Y 连锁方式，还有两种方式可排除为什么？(2)此家系最可能的遗传方式是什么？为什么？

3.17　苯丙酮尿症是一种苯丙氨酸代谢缺陷的遗传病，谱系图如下，试回答：(1)此遗传病的遗传方式？(2)谱系中哪些人可确定为是此病相关基因的杂合体？(3)Ⅲ - 2 是携带者的几率是多少？



习题 3.16 图



习题 3.17 图

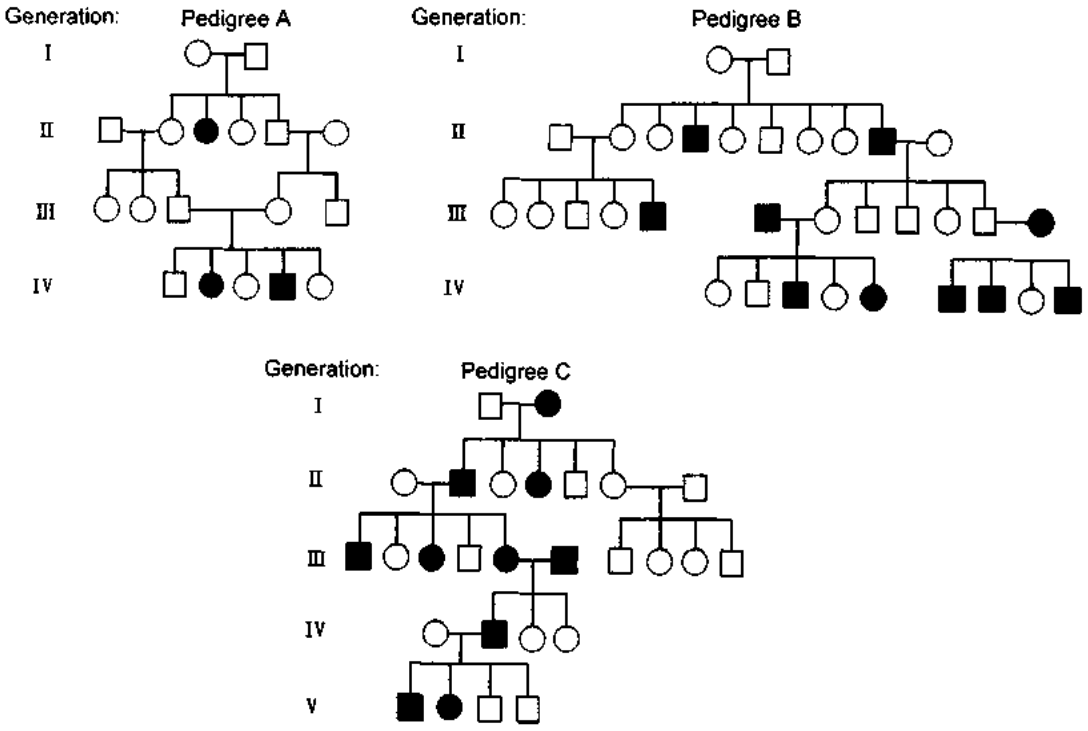
3.18　如下图谱系所示，试判断此遗传病的遗传方式。

3.19　已知一些女子是血友病(X 连锁隐性遗传)基因携带者。研究她们的凝血时间长短，结果发现个体间有很大不同；一部分人与正常人凝血时间相同，一部分与血友病患者相同，试解释。

3.20　某两栖动物在其受精卵发育的水中放入激素可使之发生性反转。当受精卵放在含有雌激素的水中，不论基因型如何，均发育为雌性表型，反之可形成表型是雄性的个体，在实验Ⅰ中，一些在含雌激素水中发育的表型雌性个体与正常雄性个体杂交。其中一半这种雌性个体只产生雄性后代，另一半产生 1∶1 的雌雄性后代。在实验Ⅱ中，经处理得到的表型雄性个体与正常雌性个体杂交，其中一半表型雄性产生 1∶1 的雌雄个体，另一半产生

3：1的雌雄个体。(1)此物种如何进行性别决定，试解释？(2)如果实验Ⅱ中的3：1的雌雄个体分别与正常个体杂交，有多少种杂交方式，每种方式的两性比例是多少？

3.21 试述伴性遗传，限性遗传和从性遗传的差别。



习题3.18图

第四章习题

4.1 小鼠中，C 为野鼠型毛色，C^h 为灰鼠型毛色，c^j 为喜玛拉雅型，c 为白化型，四个等位基因构成一个等位基因系列。其中 C 显性于其他三种。c 隐性于其他三种。C^h 显性于 c^j，试判定下列各种杂交的后代的表型。

(1) $C/C \times c/c$；　(2) $C/C^h \times C/c$　(3) $C/c \times C/c$；　(4) $C/C^j \times C^j/c$；　(5) $C/C^j \times c/c$　(6) $C^h/C^j \times C^j/c$　(7) $C^j/c \times c/c$　(8) $C/C^j \times C/c$　(9) $C/C^j \times C/C^h$

4.2 在某二倍体群体中，一特定基因有且仅有3种等位基因(w，w1，w2)，则群体中可能有多少种二倍体的基因型，试写出。

4.3 ABO血型的遗传学基础是

(1)复等位基因　(2)多表达的半合子(polyexpressivehemisyggotes)　(3)3个独立分布的基因d，等位排斥选择

4.4 人类的 I^A，I^B，i 组成决定血型的多等位基因系列，下列叙述哪些是对的，并解释。

(1)2个A型血的人可生出O型后代。(2)一个A型，一个B型的人可生出O型后代。(2)一个A型，一个O型的人可生出AB型后代。(4)一个AB型，一个A型的人可生出O型后代。(5)一个AB型，一个B型的人可生出A型后代。

4.5 一男子血型是O，M型，女子为A，M型，她的孩子是A，MN，则上述男子不可能是小孩的父亲，因为：(1)O型男子不可能有A型孩子。(2)O型男子不可能有MN型孩子。(3)O型男子和A型女子不可能有A型孩子。(4)M型男子和M型女子不可能有MN型孩子。

4.6 金鱼草中，红花(R)不完全显性于白色(r)，R/r 的杂合子为粉红色。现有一红花金鱼草与一白花金鱼草杂交，判断下列花色：(1)F_1；　(2)F_2；　(3)F_1 与红色亲本回交的后代；　(4)F_1 与白色亲本回交的后代。

4.7 在某种牛中，红色毛色基因(R)和白色毛色基因(r)的杂合体是花色。若两只花色牛杂交，其后代颜色为花色的比例为多少？

4.8　若花色牛分别与(1)红色(2)花色(3)白色牛杂交则后代情况如何?

4.9　桃子中,毛状表皮(*F*)显性于光滑表皮(*f*),卵形脐基因 *O* 和无脐基因 *o* 的杂合体产生圆形脐。现有一纯合的毛状、无脐品种与另一光滑、卵形脐品种杂交,则

(1)F_1 表型　(2)F_2 表型　(3)F_1 与光滑卵脐亲本回交产生后代的表型将如何?

4.10　豚鼠中,短毛 *L* 显性于长毛 *l*,黄色毛色基因 W 和白色基因 w 的杂合体呈花色。

有一短毛、花色豚鼠与一长毛白色豚鼠杂交,生出花色小鼠,此小鼠又和一短毛花色亲本回交,则后代情况如何?

4.11　萝卜的形状为长形(*L/L*),卵形(*L/l*),圆形(*l/l*),其颜色为红色(*R/R*)、紫色(*R/r*),白色(*r/r*)。若一长形红色萝卜与一圆形白色萝卜杂交,F_1,F_2 的情况如何?

4.12　家禽中,玫瑰状鸡冠基因(R)与豆状鸡冠基因(P)共存时产生胡桃状鸡冠,以上两基因的隐性等位基因以纯合体形式共存,则产生单一状鸡冠,试判断下列各情况的鸡冠形状:

(1)*R/RP/p* × *r/rP/p*　(2)*r/rP/P* × *R/rP/p*　(3)*R/rp/p* × *r/rP/p*　(4)*R/rP/p* × *R/rP/p*

(5)*R/rp/p* × *R/rp/p*

4.13　试判断下列各组中亲本的基因型:(1)胡桃冠与单片冠杂交产生 1/4 胡桃冠,1/4 玫瑰冠,1/4 豌豆冠,1/4 单片冠(2)玫瑰冠与胡桃冠杂交产生 3/8 胡桃冠,3/8 玫瑰冠,1/8 豌豆冠,1/8 单片冠(3)玫瑰冠与豌豆冠杂交产生 5 只胡桃冠,6 只玫瑰冠(4)胡桃冠与胡桃冠杂交产生 2 只胡桃冠,1 只玫瑰冠,1 只单片冠。

4.14　在鸡中,鸡腿有毛(*F*)显性于无毛(*f*),白色羽毛(*I*)显性于黑色(*i*)。(1)现有一腿上有毛,白色,玫瑰冠的鸡与另一无毛,白色,胡桃状冠鸡杂交,产生 2 只有毛,白色,玫瑰冠的鸡,4 只光滑,白色,胡桃冠的鸡,3 只有毛,黑色,豆状冠的鸡,1 只光滑,黑色,单一状冠的鸡,则亲本的基因型是什么?(2)腿上有毛,白色,胡桃冠的鸡,与一光滑,白色,豆状冠的鸡杂交,产生一只光滑黑色,单片冠的后代,若此杂交继续下去,则后代中表型与亲本类型相似的比例为多少?

4.15　F_2 代植株中有 3/16 是有色的,7/16 是无色的,从 F2 代有色个体中随机挑选出一株进行自交,则其后代中不发生分离的机率是多少?

4.16　果蝇基因 *L* 是隐性的性连锁基因,当它纯合时或半纯合时(在雄蝇中)是致死的。若一 *L/l* 的雌蝇与一正常雄蝇杂交,则在两个成活的后代中雄蝇的机率为多少?

4.17　一基因位点与老鼠的色素生成有关,当此位点杂合的亲本杂交,产生的后代有 3/4 是有色的,1/4 是白化的,还有一位点与毛色有关:即当两只黄色鼠杂交产生的后代中 2/3 是黄色 1/3 是灰色,当存在独立分配的灰色位点时,小鼠不会产生白化。

(1)当黄色鼠与白化鼠杂交,产生 1/2 白化鼠,1/3 黄色鼠,1/6 灰色鼠,则亲本的基因型是什么?(2)F_1 代的黄鼠相互杂交,则后代的表型比为多少?后代中黄色鼠中纯种比例为多少?

4.18　在某植物中,基因 Y 和 R 决定花色,两者都是不完全显性而相互作用产生 7 种花色:*Y/ YR/ R* = 深红、*Y/ yR/ R* = 紫红、*Y/ YR/ r* = 桔红、*Y/ yR/ r* = 玫红、*Y/ Yr/ r* = 黄色、*Y/ yr/ r* = 淡黄、*y/ yR/ R*、*y/ yR/ r*、*y/ yr/ r* = 白色

(1)当深红植株与白色(y/yr/r)杂交,F_1 和 F_2 的情况?(2)F_1 与深红亲本回交产生后代的情况?(3)当桔红植株与淡黄植株杂交,产生后代的花色?(4)当黄色植株与白色植株杂交,产生后代的花色?

4.19　(接上题):两株植株杂交,产生如下后代:1/8 深红,1/8 桔红,1/4 紫红色,1/4 玫瑰红,1/4 白色,试判断亲本的花色及基因型。

第五章习题

5.1　*a*⁺ *a*⁺ *b*⁺ *b*⁺ 与 *aabb* 杂交产生表型为 *a*⁺ *b*⁺ 的 F1 代。F_2 代表型为:a⁺ b⁺ 110, a⁺ b16, ab⁺　19, ab15(总数 160),基因 *ab* 是连锁还是独立分配?为什么?

5.2　玉米中一隐性基因 *a* 和 *b* 的双位点杂合体进行测交。后代表型 AB122,Ab118,aB81,ab79,用检验判断 ab 基因是独立分配的吗?解释理论值与实验值的偏差,你怎样证明你的解释?

5.3　为什么作果蝇染色体图谱要用杂合雌蝇和隐性纯合雄蝇。

5.4　AB/AB 与 ab/ab 杂交产生 F_1,F_1 测交产生如下表型:AB308、Ab290、ab92、aB210,则基因 *ab* 间重组率为多少?

5.5　野兔毛色中有白点显性于无白点,短毛显性于长毛。当纯合白点短毛野兔与无白点长毛野兔杂交产生 F_1。F_1 与无白点长毛亲本回交,产生如下后代:白点短毛 72,无白点长毛 69,白点长毛 11,无白点短毛 6,则毛色基因与长毛基因的图距为多少?

5.6　果蝇中,一突变体(b)是黑色,野生型是灰色,突变体(vg)是残翅,野生型是长翅。黑色长翅纯合体与灰色残

翅纯合体产生 F_1。F_1 雌蝇与黑色残翅雄蝇杂交，产生灰色长翅 283，灰色残翅 1294，黑色长翅 1418，黑色残翅 241。试计算黑色与残翅基因之间的图距。

5.7 果蝇中，长翅基因位于染色体 2，隐性基因 vg 纯合时产生残翅，等位基因 vg^+ 产生长翅。隐性基因 m 纯合时产生褐色眼，而等位基因 m^+ 产生红色眼，基因的位置不知。现把纯合褐色雌蝇与纯合残翅雄蝇杂交，产生的 F_1 代为野生型。F_1 代间相互杂交，在 6 只 F_2 果蝇中有野生型 4 只，残翅红眼 1 只，残翅褐眼 1 只。这时可推断(1)褐眼基因不是 X－连锁(2)褐眼基因不与残翅基因连锁。你还能得出什么结论？m 基因位于何染色体上？(假定此时无交换产生)。

5.8 应用下面的两点重组数据定位相关基因，给出各基因之间顺序及最小间距。

基因位点	重组率	基因位点	重组率
a, b	50	b, d	13
a, c	15	b, e	50
a, d	38	c, d	50
a, e	8	c, e	7
b, c	50	d, e	45

5.9 基因 a，b 是相互连锁的，其重组北为 10%，现在有 $ab^+/a^+b^+ \times ab/ab$，则后代中的表型及分离比各为多少？

5.10 基因 ab 是果蝇中的 X－连锁基因，两者相距 7 个单位，现有一 a^+b/ab^+ 雌蝇与野生型 a^+b^+ 雄蝇杂交：(1)后代中雄蝇的表型是 a^+b^+ 或 ab^+ 的几率是多少？(2)后代中雌蝇的表型是 a^+b^+ 的几率是多少？

5.11 果蝇中 a，b 是连锁的常染色体基因，在雌性个体中重组率为 5%，c，d 是 X－连锁的基因，相距 10 个图单位。一纯合显性雌蝇与一纯合隐性雄蝇杂交，产生的雌蝇进行测交，在测交的后代能观察到下列哪些现象：(1)雄，雌比例不同；(2)$a^+bc^+d^+$，a^+bcd，$ab^+c^+d^+$，ab^+cd 四种类型出现的机率相同；(3)考虑到杂交中还有其他基因，一些基因会发生独立分离；(4)双交换的类型比预料的要少，因为在两个区域间有干涉。

5.12 玉米中显性基因 A，C 都参与决定种子的颜色，不考虑第二个基因位点，纯合隐性植株产生无色种子。基因 A 和 C 是独立分配的，而隐性突变基因白色胚乳基因(wx)与 C 基因连锁，重组率为 20%，等位基因 Wx 产生淀粉胚乳(starchy endosperm)

(1)当 cWx/CwxA/A 测交，其后代的表型及比例为多少？(2)当 cWx/Cwx A/a 测交，其后代的表型及比例？

5.13 假定基因 a，b 是相互连锁，且发生交换的比率是 20%。(1)当 AB/AB 个体与 ab/ab 个体杂交，F_1 的基因型如何？F_1 会产生何种配子，比例是多少？若 F_1 测交，则后代中基因型及各自比例如何？(2)若 $Ab/Ab \times aB/aB$，F_1 产生何种配子，比例如何？F_1 测交，后代的基因型及比例如何？

5.14 番茄中，高茎杆显性于矮茎杆，茎杆高度与果实形状基因连锁，重组率为 20%。一高杆球形番茄与一矮杆梨形番茄杂交，其后代是：高杆球形 81，矮杆梨形 79，高杆梨形 22，矮杆球形 17。另一高杆球形与一矮杆梨形杂交，产生高杆梨形 21，矮杆球形 18，高杆球型 5，矮杆梨形 4。则两株高杆球形植株的基因型如何？若它俩杂交，其后代情况如何？

5.15 基因 a，b 位于同一染色体，相距 20 个单位，c，d 同位于另一染色体，相距 10 个单位。基因 e，f 同位于第三个染色体，现有纯合 $ABCDEF$ 与 $abcdef$ 个体杂交，产生的 F_1 与 $abcdef$ 亲本回交，则后代中如下表型的机率为多少？

(1)ABCDEF (2)ABCdef (3)AbcDEF (4)abCdef (5)abcDeF

5.16 某物种中，基因 A 显性于 a，B 显性于 b，C 显性于 c。现 $AbbCC \times aaBBcc$，产生的 F_1 进行测交，产生 4000 个后代并有多种表型。已知三种基因均位于常染色体，且 a，b 连锁，图距为 20 个单位，ab 不与 c 连锁。(1)预测的后代 4000 个可能出现的表型及比例。(2)当 a，b 完全连锁时，哪些表型不出现？(3)当 a，b 不连锁时，哪些表型不出现？(4)假设 a，b 不连锁，F_1 的自交后代的表型及比例？

5.17 三个基因连锁，可能有几种线性排列顺序？当一个位点于染色体最右端，还有几种排列顺序？

5.18 在酿酒酵母、红色面包霉、衣藻的减数分裂中，发生何种事件以产生亲二型、非亲二型、四型？

5.19 红色面包霉中有如下杂交，试判断基因 a，b，c 的顺序及图距。

$a^+b \times ab^+ \rightarrow 981$	$a^+c \times ac^+ \rightarrow 850$	$b^+c \times bc^+ \rightarrow 850$
1000 ab^+	853 ac^+	850 bc^+
10 a^+b^+	169 a^+c^+	140 b^+c^+
9 ab	148 ac	160 ac
2000	2000	2000

5.20 四种不同的白化红色面包霉分别与野生型杂交，均产生一半野生型，一半白化型。然后 1 号菌株又与其他

三种菌株杂交，结果如下。则哪些突变体代表不同的基因，哪些基因是连锁的，为什么？

杂交结果：1 × 2→975 白化型，25 野生型；　　1 × 3→1000 白化型；

1 × 4→750 白化型，250 野生型。

5.21　基因 *met* 和 *thi* 在红色面包霉中连锁，为定位 *arg*，使 *arg* + + × + *thimet*，得到如下子囊孢子，试作图。

arg	*thi*	*met*	26	*arg*	+	+	51
arg	*thi*	+	17	+	*thi*	+	4
arg	+	*met*	3	+	+	*met*	14
+	*thi*	*met*	56	+	+	+	29

5.22　现有一 mt^+ 成熟型，粉色（p^-）酵母株与一 mt^+ 成熟型，灰色（p^+）酵母菌株杂交，产生如下的四分体，试问 *p* 和 *mt* 是否在同一染色体上？

四分体的数	四分体的种类
18	$p^-\,mt^+$，$p^+\,mt^+$，$p^-\,mt^-$，$p^-\,mt^-$
8	$p^+\,mt^+$，$p^-\,mt^+$，$p^+\,mt^-$，$p^-\,mt^-$
20	$p^+\,mt^-$，$p^+\,mt^-$，$p^-\,mt^+$，$p^-\,mt^+$

5.24　在酵母中有如下杂交（*leu* + × + *rib*）产生如下子囊（此时基因顺序均为假设）试作出连锁图谱，并确定图距。

110		45		6		45	
leu	+	*leu*	*rib*	+	+	*leu*	+
+	*rib*	*leu*	+	*leu*	*rib*	+	*rib*
leu	+	+	+	*leu*	*rib*	+	+
+	*rib*	+	*rib*	+	+	*leu*	*rib*

5.25　红色面包霉中一染色体的一个臂上有 a、b、c 三个基因，现 ab + × + + c 得到如下子囊，试判断基因顺序，并计算所有基因间及基因一着丝粒间的距离。

45	5	146	1	10	20	15	58
ab +	ab +	ab +	ab +	ab +	ab +	ab +	ab +
+ bc	a + +	ab +	+ + +	a + c	+ + c	abc	+ b +
a + +	+ bc	+ + c	abc	+ b +	ab +	+ + +	a + c
+ + c	+ + c	+ + c	+ + c	+ + c	+ + c	+ + c	+ + c

5.26　若在红色面包霉中，一个两点杂交的后代中有 10% 子囊是重组的，则此两点间图距是多少？

5.27　一个 $met^-\,trp^-$ 营养缺陷型菌株与一个 $leu^-\,nic^-$ 营养缺陷型形成异核体，从异核体中选出二倍体，在此二倍体的准必性世代里，从分生孢子中得到如下 8 个单倍体菌株，且出现频率几乎一致，则此四个基因哪些是连锁，哪些是不连锁的？

1.	nic^+	leu^+	met^-	trp^-	5.	leu^+	nic^+	met^+	trp^+
2.	met^+	leu^+	nic^+	trp^+	6.	met^-	nic^-	leu^-	trp^-
3.	trp^+	met^-	leu^+	nic^-	7.	trp^+	leu^-	met^+	nic^+
4.	leu^-	nic^-	trp^+	met^+	8.	nic^+	met^-	trp^-	leu^-

5.28　绿色二倍体菌株对下列隐性突性基因都是杂合的，*sm*、*pu*、*phe*、*bi*、*w*（白色）、*y*（黄色）、*ad*。从此二倍体中得到黄色，白色两组单倍体，其基因型如下。则这些基因分布在几个连锁群中，每个基因分别位于哪个连锁群？

基因型

sm	*pu*	*phe*	*bi*	*w*	*y*	*ad*
sm	*pu*	*phe*	+	*w*	*y*	*ad*
+	*pu*	+	*bi*	*w*	+	*ad*
+	*pu*	+	*bi*	*w*	+	*ad*
sm	+	*phe*	+	+	*y*	+
+	+	+	+	+	*y*	+
sm	*pu*	*phe*	*bi*	*w*	+	*ad*

第六章习题

6.1 下表中为实际量度的鸭子的头宽和翅长，(1)求表中8只鸭子的头宽和翅长的平均值及标准方差；(2)求头宽和翅长的相关系数；(3)用什么方法来计算就能得知头宽和翅长之间的联系。

编号	头宽(cm)	翅长
1	2.75	30.3
2	3.20	36.2
3	2.86	31.4
4	3.24	35.7
5	3.16	33.4
6	3.32	34.8
7	2.52	27.2
8	4.16	52.7

6.2 为什么两个纯种亲本杂交产生的 F_1 代之间的差异并不比每个亲本之间的差异大？

6.3 两对基因 *A/a* 和 *B/b* 在群体中决定根累加的高度，纯合子 *AABB* 高 50cm，*aabb* 高 30cm(1)二者杂交的 F_1 代高多少？(2)F_2 代高为 40cm 的基因型应是怎样的？(3)F_2 代中 40cm 高的植株频率可能是多少？

6.4 三个独立分配的基因(*A*、*B*、*C*)，每个座位有两个等位基因决定植物的高度。每个等位基因中大写字母要比基本高度 2cm 加 2cm。(1)两个纯合子 *AABBCC*(14cm) × *aabbcc*(2cm)之间杂交代 F_1 的高度将是多少？(2)$F_1 \times F_1$ 杂交 F_2 的高度(频率和表型)的分布如何？(3)在 F_2 代中有多少比例的后代高度和原来两个亲本(*AABBCC* 和 *aabbcc*)相等。(4)F_2 中多少比例的后代其高度和 F_1 相同？

6.5 假设在旱金莲中由三对基因控制花的长度。这三对基因作用相等并具有相加效应。一个完全纯合的亲本花长 10mm，与另一个花长 30mm 的纯合亲本杂交，F_1 植物的花长为 20mm 长。F_2 植物的花长在 10 到 30mm 范围内。F_2 中花长 10mm 和 30mm 各占 1/64。你预计在 F_1 代植物与花长 30mm 亲本回交的后代中花长分布将如何？

6.6 一种大麦(*asplund*)的节间距平均长度 2.12mm，而在另一品种中(*abed binder*)节间距平均长度为 3.17mm，将二者杂交 F_1 的节距为 2.7mm，F_2 的节间距变异范围在 2.12～3.17mm 之间。分析 F_2 代表明 F_2 代的 125 株中有 8 株为 *asp/and* 型，节距平均为 2.19mm，另外 8 株为 abed binder 型，平均节距 3.24mm，问大麦节距这一性状是否为数量性状？为什么？

6.7 在上题中根据已知信息推测节距涉及几对基因？

6.8 假设两种不同类型的燕麦，一种每株产量为 4g，另一类型为 10g，此性状由 3 对等效积加多基因控制的(*AA*、*BB*、*CC*)。若将以上两种类型杂交，F_1 和 F_2 的表现如何？它们的分布如何？

6.9 两个不同品系的植物 A 和 B，A 品系的蛋白质含量高于 B 品系，而 B 品系的抗寒性状高于 A 品系。二个品系之间有 10 个基因座的差异，没有显隐性关系，若二者杂交，F_2 代中有多少比例的个体完全具备 AB 品系的优良性状？

6.10 在蚕豆的群体中，种子平均重 0.88g。已知种子重是受到加性方差的影响。一个农民选择了两种种子，重 1.02g 他将这两种蚕豆种种下去，并进行杂交，收集了后代种子，平均重为 0.96g。问种子重量的狭义遗传力是多少？

6.11 大鼠的 SHR 品系对盐敏感，在高盐环境下生长会出现高血压。而另一品系 TIS 品系对盐不敏感。若将 SHR 品系的大鼠放在一个含盐量不同的环境中，有的大鼠比另一些大鼠的血压高，在此群体中血压的遗传力将是怎样的？

6.12 一位学者想确定小鼠尾长的狭义遗传力，他测量了群体中小鼠的尾长，计算其平均长度为 9.7cm。然后他在群体中选择了 10 只尾巴最长的小鼠，这 10 只小鼠的平均尾长为 14.3cm。他将这些鼠进行杂交，并检测下一代的尾长，F_1 的平均尾长为 13cm。请计算选择的差异和小鼠的狭义遗传力。

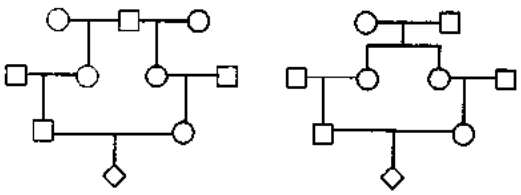
6.13 9 名大学生的身长及他们父亲的身高如表所示：(1)分别计算孩子们和父亲身高的平均值；(2)计算二者的相关系数；(3)通过儿子高度对父亲高度的回归来确定狭义遗传力。

儿子高度(英吋)	父亲高度(英吋)
70	70
72	76
71	72

64	70
66	70
70	68
74	78
70	74
73	69

6.14　一个品种的鸡其卵重的狭义遗传力是 0.60，而农民从这群鸡中进行选择以期增加卵重，选择组和非选择组的平均卵重相差 10g，选择组下一代卵重将增加多少？

6.15　家系如图的婚配，求近亲系数。



习题 6.15 图

第七章习题

7.1　表中所示酶的活性区域和人类两种先天代谢缺陷有关(隐性遗传)。很多代谢遗传病有着相似的情况。(1)用这些信息可以进行遗传咨询吗？(2)对病人的研究能清楚地说明病因吗？(3)从资料中请重新评价显性的概念。

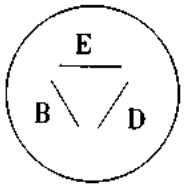
疾病	有关的酶	酶的活性区域		
		患者	患者的双亲	正常人
过氧化物酶缺陷	过氧化物酶	0	1.2～2.7	4.3～6.2
半乳糖血症	Gal－1－p 尿苷转移酶	0～6	9～30	25～40

7.2　两个白化人结婚，生育了一个正常的孩子，请解释可能的原因。(至少提出两种以上可能原因)

7.3　人类的苯丙酮尿症(PKU)是由于下列生化反应中 A 阶段酶的失活所致，而尿黑酸症(AKU)是 B 阶段酶的失活所致，一个 PKU 患者和一个 AKU 患者结婚，你预计他们的孩子表型如何？

$$\text{苯丙氨酸} \xrightarrow{A} \text{酪氨酸} \xrightarrow{A} CO_2 + H_2O$$

7.4　有三个单独分离的酵母品系 *trpB*，*trpD*，和 *trpE*，在有充足的 Trp 平板上，它们可以生长，但长的较弱。现将三个品系的菌液在此平板上划线，为了不使它们接触，将线划成三角形，结果 *trpE* 线两端和 *trpD* 线一端生长良好。(1)你认为这和互补有关吗？(2)请解释生长好的原因；(3)从实验中能推断 Trp 合成中的步骤和酶缺陷？(4)为了论证这样的生长方式为什么要在培养基上加少量的 Trp？

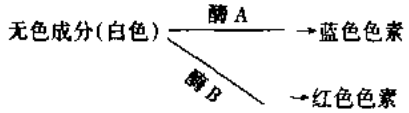


成　分　测　验

	A	B	C	D	E	G
1	—	—	—	+	—	+
2	—	+	—	+	—	+
3	—	—	—	—	—	+
4	—	+	+	+	—	+
5	+	+	+	+	—	+

7.5　分离了一系列突变，它们的生长都需要加 G。已知在 G 的生化合成途经中的成分是 A、B、C、D、E、G 但不知其顺序。经测验每一成分能通过突变体(1～5)的生长来加以推测。在下表中"＋"表示加入生长，"－"表示加入不生长。(1)A～E 的顺续如何排列？(2)在此途径的什么位置各种突变体被阻断？(3)由 1、3 和 2、4 组成的异核体在基本培养基上能生长吗？1、3 和 3、4 呢，1、2；2、4 及 1、4 呢？

7.6 在一种植物中，花瓣是紫色的。它有两种突变体，一种*(m1)*纯合时，*(m1m1)*产生蓝色花瓣，另一种突变体*(m2)*纯合时*(m2m2)*开红色花。生物学家研究此植物花中色素的合成，已描述了下列合成途径：



(1)你预期突变体A酶缺乏活性，表型如何？(2)基因型 *M1m1*、*M2m2* 植物的花表型如何？(3)如 *M1m1M2m2* 回交，你预计后代花的颜色和比例如何？(4)为什么这种花的突变是隐性的。

7.7 下图表示四种缺失(1～4)都涉及到T4噬菌体的rⅡA顺反子。rⅡA中5个位点(a～e)突变有的能抗缺失突变(+)，有的不能抗缺失突变(−)，产生r⁺重组。(1)突变顺序是怎样的？(2)另一品系的T4噬菌体在rB有突变位点。将此品系依次与每个rⅡA缺陷品系混合，再去感染 *E. coli* K使每个寄主细胞至少被一个rⅡA和一个rⅡB突变体感染。正常的噬菌斑带有缺失1，2和3，没有带有4的。若B顺反子在A的右侧，请释缺失者的行为。

	a	b	c	d	e
1	+	+	−	+	+
2	+	+	−	−	−
3	−	−	+	−	+
4	+	−	+	+	+

习题 7.7 图

7.8 在链孢霉的单倍体中，假设两个 *Val* 合成特定基因。它们的突变等位基因是 *Val-1* 和 *Val-2*。而野生行的等位基因是 *Val-1⁺* 和 *Val-2⁺*。这两个基因连锁在同一条染色体上。交换发生在两者之间，平均两次减数分裂中就有一次发生交换。(1)两个基因之间未发生交换的减数分裂的比例是多少？(2)两个基因之间的重组频率是多少(以图距单位表示)？(3) *Val-1Val-2⁺* × *Val-1⁺ Val-2* 杂交后代放在 *Val* 的平板上，有多少比例的后代能生长？(4) *Val-1*　*Val-2⁺* 品系积聚的中间成分B，而 *Val-1⁺*　*Val-2* 品系积聚的中间成分是A。*Val-1 Val-2* 品系在有Val或A的平板上生长，但 *Val-1⁺ Val-2* 品系仅在有 *Val* 平板上生长，不能在有B的平板上生长。说明Val的合成途径，及基因控制的位置。

7.9 在噬菌体中，将一套缺失突变进行互补配对互交，结果如右表，(+号表示野生型重组体)从表中推出缺失的图谱。

7.10 在人类中某些基因已知具有表型的多效性，那么此是否可以否定一基因一酶学说呢？请解释？

7.11 在柯林西草(*Collinsia parviflora*)中花瓣的正常颜色是紫的。有4种隐性突变每种的花瓣都是白的。四种突变品系互交重组，结果如表中所示：(1)请解释结果，并用图表示过程。(2)若将2*3和2*4杂交，它们的 F_1 和 F_2 将如何？

	1	2	3	4	5
1	−	+	−	+	−
2	+	−	+	+	−
3	−	+	−	−	−
4	+	+	−	−	+
5	−	−	−	+	−

习题 7.9 图

突变体杂交	F_1	F_2	
1×2	全紫	1/2 紫	1/2 白
1×3	全紫	9/16 紫	7/16 白
1×4	全白	全白	全白

习题表 7-11

7.12 一位学者对丝状真菌曲霉(*Aspergillus*)单倍体中Leu的合成感兴趣。他用诱变剂处理孢子而获得35种点突变(a～e)，都是营养缺陷型。他首先制备突变体之间的异核体来检查这些位点的相关功能。"+"号表示异核体生长，"−"号表示不生长。在培养基上加Leu，结果如表所示。然后他有将突变体杂交，每种杂交都测定500个子囊孢子，即将孢子放在缺乏Leu的培养基上培养，结果如下(表中的数字为500个后代中原养型的数。请解释两套

	a	b	c	d	e
a	−	+	+	+	−
b		−	+	+	+
c			−	+	+
d				−	+
e					+

	a	b	c	d	e
a	0	125	128	126	0
b		0	124	2	125
c			0	124	127
d				0	123
e					0

习题表 7-12

结果(注意 b－d 杂交而形成的 Leu 原养型不是因为回交突变所致)。

第八章习题

8.1 若一个营养缺陷型 *E. coli A* 只能生长在含有 T 的培养基上，另一个营养缺陷型 *B* 能生长在含有 Thr 的培养基上，你将怎样测定 A 品系的 DNA 是否能转化 B 品系？

8.2 *E. coli* 的 F^-，F^+，F' 和 *Hfr* 品系之间的差别是什么？

8.3 在 $F' \times F^-$ 的杂交中，F^- 受体以很高的频率变成为供体，然而在 $Hfr \times F^-$ 中很少有受体变成了供体，解释为什么？

8.4 用中断杂交技术对四个品系的 *Hfr* 品系的顺序进行检测，他们将一定量不同的基因传递到 F^- 受体中。每一 *Hfr* 品系传递的基因的顺序都各不相同。(如表所示)。问原来的基因顺序是怎样的？*Hfr* 四个品系的原点和方向如何？

Hfr 品系	传递顺序
1	Q－F－B－A－E－M
2	R－H－M－E－A－B
3	E－M－H－R－C－X
4	O－G－X－C－R－H

8.5 当 *Hfr* 供体和 F^- 受体结合时，若 F^- 受体带有原噬菌体，受体可存活。若供体 *Hfr* 带有原噬菌体，而受体 F^- 不带时合子通常在 100 分钟后会裂解，此称为合子诱导。

(1)解释合子诱导；　(2)怎样检测原噬菌体整合的位置？

8.6 溶源周期和裂解周期之间的区别是什么？

8.7 普遍性转导和特异性转导之间的区别是什么？

8.8 指出下列的 9 种现象是属于何种转导？普遍性转导(GT)；特异性转导(ST)；二者兼有(B)；二者中的一个(N)(1)噬菌体带有细菌的 DNA 或本身的复制起点，或二者都没有；(2)噬菌体带有自己的 DNA 并与细菌 DNA 共价连接；(3)噬菌体整合到宿主的特异位点；(4)噬菌体随机整合到宿主的任何位点；(5)将宿主的 DNA 包装到头部；(6)宿主溶原化；(7)原噬菌体状态存在；(8)和温和噬菌体有关；(9)和烈性噬菌体有关。

8.9 分析下列转导资料：问 *dhl* 和 *leu* 那一个和 *acef* 紧密连锁？

8.10 分析下列 P1 转导资料，问 trpE 或 trpB 那一个和 cysB 紧密连锁。

8.11 根据下列共转导资料确定 *cheA*, *cheB*, *eda* 和 *supD* 的顺序。

供体	受体	选择标记	非选择标记	%
$acef^+$ del	$acef^-$ dhl^+	$acef^+$	dhl	88
$acef^+$ leu	$acef^-$ leu^+	$acef^+$	leu	34

供体	受体	选择标记	非选择标记	%
$cysB^+$	cyBtrpE$^+$	$cysB^+$	trpE	37
$cysB^+$ trpB	cyBtrpB$^+$	$cysB^+$	trpE	53

标记	共转导频率%
cheA – *eda*	15
cheA – *supD*	5
cheB – *eda*	28
cheB – *supD*	2.7
eda – *supD*	0

8.12 将 T4 噬菌体原液稀释到 10^{-8}，吸 0.1ml 10^{-8} *E. coli* B/ml 的菌液及 2.5ml 的已溶解的琼脂，涂布在一个琼脂平板的表面上，次日出现了 20 个噬菌斑。问原来 T4 噬菌体的浓度是多少？

8.13 下面是用二因子杂交来分析噬菌体 λ4 个基因(*c*、*mi*、*s* 和 *co*)之间的连锁，根据资料画出四个基因的连锁图。

亲本	后代											
c+ × + mi ⟶	1213	c	+ ,	1205	+	mi,	84	+	+ ,	75	c	mi
c+ × + s ⟶	566	c	+ ,	808	+	s,	19	+	+ ,	20	c	s
co+ × + mi ⟶	5162	co	+ ,	6510	+	mi,	311	+	+ ,	341	co	mi
mi+ × + s ⟶	502	mi	+ ,	647	+	s,	65	+	+ ,	56	mi	s

8.14 噬菌体 T4 的 r Ⅱ突变型能在 *E.coli* B 菌株中生长，但不能在 *E.coli K12* 中生长。将两个 *r* Ⅱ突变型双重感染 *E.coli* B，将裂解稀释 6×10^{7} 倍接种于 *E.coli* B 中，将裂解物稀释 2×10^{5} 倍接种于 *E.coli* K12(λ)，结果在 K12(λ)品系中产生 12 个噬菌斑，而在 B 菌株上产生了 16 个噬菌斑，计算这两种突变型之间的重组数。

8.15 在果蝇中突变体 A，B，C，D，E，F 和 G，都有相同的表型：眼中缺少红色素，在互补测验中，逐对组合的结果如下："+"表示可以互补，"－"表示不可以互补，问 ABCDEFG(1)这些突变在几个基因中？(2)那些突变在同一个基因中？

	A	B	C	D	E	F	G
G	+	−	+	+	+	+	−
F	−	+	+	−	+	−	
E	+	+	−	+	−		
D	−	+	+	−			
C	+	+	−				
B	+	−					
A	−						

第九章习题

9.1 在 Griffith 发现转化以后，Avery、Macleod 和 McCarty 研究了相同的现象，(1) 他们的实验是怎样设计的，比 Griffith 的实验有何发展？(2)如和用酶来控制实验？

9.2 Hershey 和 Chase 通过不同的同位素来标计 T2 的外壳蛋白和 DNA 得出的结论是：(1) 只有蛋白质进入了被感染细胞，(2) 整个的噬菌体都进入了被感染细胞；(3) 中期染色体是由两条染色单体所组成，每条单体含有单个的 DNA 分子；(4)噬菌体的遗传物质可能是 DNA；(5)噬菌体的外壳蛋白直接合成子代噬菌体。

9.3 有什么证据表明烟草镶嵌病毒(TMV)的遗传物质是 RNA？

9.4 在 DNA 和 RNA 中糖分子的那一个碳原子和磷酸键相连接？

9.5 什么碱基对 DNA 是特有的？什么碱基对 RNA 是特有的？核苷酸和核苷有什么不同？

9.6 有什么证据表明 DNA 双螺旋中一对碱基是有一个嘌呤和一个嘧啶组成。

9.7 有什么证据表明 DNA 中只有两种碱基对 A－T 和 G－C 构成？

9.8 在 DNA 分子中有多少不同种类的核苷酸？

9.9 有什么证据表明 DNA 是遗传物质？

9.10 DNA 双螺旋模型建立的基础是什么？

9.11 什么是 Z－DNA？其特点是什么？有何生物学意义？

9.12 从细菌的噬菌体中分离了 DNA，经鉴定此 DNA 含 33%A，26%T，18%G、和 23%C，从此资料中可以推测此噬菌体的 DNA 的结构如何？

9.13 什么叫酶切图谱？建立酶切图谱有几种方法？

9.14 我们希望使基因组随机切成很小的片段(250bp)左右，应当选择什么样的限制酶，若想切成 4kb 的随机片段需要选什么样的限制性酶？

9.15 限制酶有几种？结构和功能各有什么特点？

9.16 若在 106nt RNA 的核苷酸的排列是随机的，而各种碱基的比例是 20%A、25%C、25%U 和 30%G，，你预计 5′－GUUA－3′这一序列可以出现多少次？

第十章习题

10.1 原核生物基因组的结构特点是什么？和真核生物基因组有什么不同？

10.2 什么是类核，它的结构特点是什么？有何生物学意义？

10.3 在细菌中常见的质粒有几种？它们有什么特点？

10.4 什么是重叠基因，有何生物学意义？在真核细胞是否存在重叠基因？

10.5 什么是人类基因组计划？其意义何在？

10.6 常染色质和异染色质有什么差别？

10.7 核小体模型是怎样建立的？模型的结构如何？

10.8 简述染色质的侧环模型。

10.9 着丝粒的结构特点是什么？有何功能？

10.10 端粒结构特点是什么？有何生物学功能？

10.11 什么叫做核型？有何应用价值。

10.12　常见的染色体带型有几种？有何应用价值。

10.13　什么是单一顺序？编码顺序是否都是单一顺序？如何应用复性动力学曲线来区分单一顺序和重复顺序？

10.14　什么叫做假基因？什么叫做加工基因？它们的结构有什么特点？可能是怎样产生的？

10.15　基因家族和基因簇这两个概念有什么区别？

10.16　什么是短的分散因子？什么叫长的分散因子？它们是怎样产生的？

10.17　人类的 Alu 顺序结构是怎样的？有何应用价值？怎样分离 Alu 顺序？

10.18　什么叫做卫星 DNA？什么叫做小卫星 DNA 和微卫星 DNA？有何应用价值？

10.19　什么叫做 C 值矛盾？你如何解释 C 值矛盾？

10.20　内含子是怎样发现的？它有何生物学意义？

10.21　有人说只有真核生物才有内含子，内含子是不编码的这种说法对吗？为什么？

10.22　从 T2DNA 分子的群体中挑出了两个片段：这些分子通过热变性链相互分离，然后又可以复性。请用图表示(1)2 和 3 复性时的情况，(2)3 和 4 复性时的情况，

1　TAGCTCC　　3　GCTCCTA

2　ATCGAGG　　4　CGAGGAT

10.23　(1)你若将 T4 染色体进行变性和复性将会产生什么样的结构？(2)若用 T7 染色体进行变性和复性，将会产生什么不同的情况？(3)若用 λ 呢？

10.24　将下例各种 DNA 和相应的染色质相匹配起来

DNA 类型	染色质类型
Barr 小体	常染色质
着丝粒	兼性异染色质
端粒	组成性异染色质
大部分表达基因	

第十一章习题

11.1　原核生物复制起点的结构有何特点？

11.2　*E.coli* 复制起始复合物有那些蛋白组成？它们的功能是什么？

11.3　DNA pol Ⅰ和 pol Ⅲ的功能有什么不同？

11.4　pol Ⅲ由哪些亚基组成的，它们的功能是什么。

11.5　什么是 DNA 的成环复制模型，或复制体模型。

11.6　什么是滚环复制？它有何特点？举出原核和真核生物滚环复制的例子。

11.7　哪些重要的实验证实了复制半保留模型，Meselson－Stahl 实验还存在什么问题？

11.8　有哪些方法可以确定原核环状 DNA 的固定的复制起始点，以及是单向复制还是双向复制？

11.9　解释用双向电泳确定离复制起点不同距离位点的顺序。

11.10　λ 的复制起始点的结构特点是什么？起始复合体中有哪些蛋白质因子？其功能是什么？

11.11　以 T7 为例说明噬菌体线性 DNA 的复制终止机制。

11.12　冈崎脉冲实验的结果表明新合成的 DNA 片段都是小片段，这和半不连续复制模型是不相吻合的，你怎样加以解释？

11.13　以 SV40 为例说明真核生物的复制是怎样起始的。

11.14　真核生物复制起始时有哪些蛋白质因子和酶介入，他们的功能是什么？

11.15　真核生物的复制特点和原核生物的复制特点有什么不同？

11.16　在 Meselson 和 Stahl 实验中，将^{15}N标记细胞移到含^{14}N的培养基中，这样我们就可标明第一代。(1)对于半保留复制模型而言，我们可预期在第 1,2,3,4,6,和 8 代中$^{15}N-^{15}N$,$^{15}N-^{14}N$,$^{14}N-^{14}N$的比例将是如何？按 DNA 复制的全保留模型，预期实验的结果将样的？

11.17　快停突变和慢停突变能帮助我们筛选出什么物质。

11.18　当用 5－溴尿嘧啶(BrdU)来处理 DNA 时，在复制时它可以掺入到染色体中，在紫外线的照射下可使被掺

人的 DNA 的一条单链破坏，而另一条单链是正常的，当双链都被 BrdU 分子插入时，经 UV 照射后都不能被吉姆萨染色，而前者仍可染色。若在细胞培养的第一周期加入 BrdU，那么周期分别取样，制各染色体中期片，经染色后会出现什么情况？

第十二章习题

12.1 转录和复制都是合成的过程，二者有何不同？
12.2 细胞 RNA 聚合酶由几个亚基组成，RNA 聚合酶有何功能？
12.3 原核生物启动子结构有什么特点？
12.4 RNA pol 核心酶和全酶在 DNA 上如何分布？
12.5 RNA pol 核心酶在 DNA 如何移动的？
12.6 转录有几种终止结构，特点如何？
12.7 真核生物的转录和原核有何不同？
12.8 真核生物Ⅱ类启动子的结构是怎样的？
12.9 增强子的调控有何特点？
12.10 Ⅰ类启动子的特点，分布及合成的产物是什么？
12.11 RNA pol Ⅰ的转录起始需要哪些蛋白质因子？它们的功能是什么？
12.12 Ⅲ类启动子结构有什么特点？在细胞中分布如何？主要转录何种 RNA？
12.13 Ⅱ类启动子转录起始需要哪些蛋白质因子？它们的功能是什么？
12.14 Ⅲ类启动子有哪些蛋白质参与起始，它们的功能是什么？

第十三章习题

13.1 原核生物转录后加工，主要包括哪些方面？
13.2 原核生物的 tRNA 是如何加工的？
13.3 原核生物的 mRNA 是否经过加工，试举例说明。
13.4 真核生物 mRNA 加工主要包括哪些步骤？
13.5 真核生物的 rRNA 是如何加工的？
13.6 真核生物的 tRNA 的加工和原核 tRNA 的加工有何不同。
13.7 真核生物 tRNA 内含子的剪切有何特点？
13.8 真核生物 tRNA 的内含子的结构有何特点？
13.9 真核生物 mRNA 的内含子有几种？其特点如何？
13.10 内含子的剪切时 5′外显子和 3′端外显子离的很远，一般用什么策略将它们拉近？
13.11 Ribozyme 是如何发现的？有何意义？
13.12 什么是转酯反应，那些生物学过程涉及转酯反应？
13.13 什么叫做编辑？有何生物学意义？
13.14 编辑的分子机制是什么。
13.15 植物病毒和类病毒 RNA 自我剪切的机制是什么？
13.16 什么是反式剪切？举例说明其分子机制。
13.17 Ribozyme 的发现对于进化有意义？

第十四章习题

14.1 遗传密码是如何破译的？
14.2 遗传密码的特点是什么？
14.3 什么是摆动假说？（是怎样产生的？）有何生物学意义？
14.4 线粒体遗传密码和通用密码有何不同？此意味着什么？
14.5 各种密码子的使用机会是否相同？有何生物学意义？

14.6　tRNA 的二维结构的特点是什么?
14.7　tRNA 的三维结构的特点是什么?
14.8　tRNA 对氨基酸如何识别?
14.9　什么是 tRNA 的 identity? 有何特点?
14.10　什么是校正基因? 有几种类型?
14.11　抑制基因是否会影响蛋白质合成的正常终止? 为什么?
14.12　细菌核糖体有哪些活性位点,其功能位置是怎样的?

第十五章习题

15.1　氨基酰 tRNA 是如何形成的?
15.2　氨基酰 tRNA 合成酶如何控制识别正确的 tRNA 和氨基酸?
15.3　什么是蛋白质合成的三位点假设。
15.4　在细菌翻译起始中,有哪些起始因子参与? 它们有何功能?
15.5　起始 RNA 有何特点?
15.6　真核生物的翻译起始和原核生物有何不同?
15.7　在原核和真核细胞中,翻译的终止有哪些因子介入,其功能是什么?
15.8　在真核生物中为什么采用单顺反子 mRNA 的策略,而不用多顺反子 mRNA?
15.9　在真核细胞中有几类不同的蛋白转运途径?
15.10　什么是信号假说?
15.11　简介信号识别蛋白的结构和功能。
15.12　转运到细胞器中的蛋白的导肽结构和功能有何特点?
15.13　细菌的蛋白跨膜机制是什么。
15.14　若我们在一外源基因前加上一段信号顺序,在 *E. coli* 中表达,是否可以使其分泌? 为什么?
15.15　蛋白质的加工有几种形式?
15.16　糖基化是怎样进行的,有何生物学意义?
15.17　什么是蛋白内含子? 它和一般的蛋白质的剪切加工有何不同?

第十六章习题

16.1　什么是操纵子模型? 操纵子的建立有何意义?
16.2　何为正调控? 何为负调控? 何为诱导? 何为阻遏?
16.3　简述乳糖操纵子中阻遏蛋白的结构和功能。
16.4　何为 CAP 结合位点? 其存在有何意义?
16.5　试述半乳糖操纵子的结构和特点。
16.6　试述阿拉伯糖的结构和特点。
16.7　色氨酸操纵子有何特点?
16.8　什么叫衰减子? 其有何作用?
16.9　衰减子存在于哪些操纵子中? 真核生物有几类似的调控形式? 为什么?
16.10　在噬菌体中采用哪些策略来调控转录起始?
16.11　举例说明在原核生物中采用 DNA 重排来调控转录起始的机制。
16.12　什么是反义调控? 在原核中存在哪些反义调控的例子。
16.13　原核生物在翻译水平上采用哪些调控策略?
16.14　*E. coli* 外膜蛋白 OmpC 和 OmpF 对不同的渗透压如何进行调控?
16.15　什么是反义技术? 有何应用价值?
16.16　稀有密码子在调控中有何作用?

第十七章习题

17.1 简述λ噬菌体的裂解周期,溶源周期及二者之间的关系。
17.2 简述λ噬菌体基因组的基本结构以及一些重要基因的功能。
17.3 简述λ噬菌体早期转录,晚期转录及其调控。
17.4 λ噬菌体启动子 P_R 和 P_L 各有什么功能,如何启动?
17.5 λ噬菌体的C Ⅱ蛋白和C Ⅲ蛋白的功能是什么?
17.6 试述C Ⅰ蛋白的结构和功能?
17.7 为什么C Ⅰ蛋白在低浓度时可以活化 P_R 启动子,而在高浓度有可抑制 P_R 启动子。
17.8 为什么紫外线照射可以诱导溶源化的λ噬菌体裂解?
17.9 Cro 蛋白的结构和功能是什么?
17.10 反终止蛋白反终止的机制是什么?
17.11 裂解或溶源化途径是如何决定的?
17.12 在λ噬菌体中有那些属于反向调控?
17.13 λ噬菌体左向转录已转录了 *int* 基因,为什么 *int* 基因本身还需要有自己的启动子?
17.14 为什么游离的λ噬菌体 P_I 启动的转录会产生 sib 位点,而 P_I 启动的转录就不会产生 sib 位点呢?当溶源状态时是否会产生 sib 位点呢?为什么?

第十八章习题

18.1 真核生物的基因调控有什么特点?
18.2 为什么真核生物中没有发现操纵子这一调控形式?
18.3 真核生物的基因簇也是连锁在一起的,它和原核的操纵子有何不同?
18.4 什么是 DNA 的扩增?在真核生物中 DNA 扩增有几种形式?
18.5 异染色质化的本质是什么?怎样用实验来证实?
18.6 什么是 DNA 的重排?试举例说明真核生物中重排对基因表达的调控作用。
18.7 真核生物在选择性加工时有几种调控途径?
18.8 举例说明在真核生物中反义 RNA 的调控。
18.9 什么叫做座位控制区?它有什么作用?
18.10 什么叫做限定子(delimiter)或绝缘体(insulator),它有什么功能?
18.11 什么叫做核基质区?在调节中起什么作用?
18.12 组蛋白和非组蛋白在真核生物基因表达调节中各起什么作用?
18.13 什么是 DNase 敏感位点,超敏感位点,和染色质结构有什么关系?
18.14 顺式作用元件有哪些类型?
18.15 反式作用因子和顺式作用元件的结合有哪几种结构模式?
18.16 mRNA 的稳定性和表达调控有什么关系。

第十九章习题

19.1 肿瘤细胞有何特点?
19.2 什么叫确定细胞系?它有何特点?
19.3 什么叫做癌基因?
19.4 癌基因按其功能可分为哪几组?其功能是什么?
19.5 什么叫做抗癌基因,常见的有哪几种?
19.6 *Rb* 基因抗癌机理是什么?
19.7 P21 蛋白在细胞周期中起到什么作用?
19.8 P27 蛋白在细胞周期中的作用是什么?

19.9　什么是P53蛋白，它有什么功能？

19.10　P53有哪些活性？

19.11　*scr*基因的产物有功能？突变后为什么会导致肿瘤的发生？

19.12　简述c-*erbB*产物的结构与功能，以及突变后致癌的机制。

19.13　简述c-*sis*产物的结构与功能，以及突变后致癌的机制。

19.14　Ras蛋白在细胞信息传递中起什么作用？其突变后会产生什么影响？

19.15　癌基因活化的机制是什么？

19.16　激发癌基因活化有哪些途径？

第二十章习题

20.1　线粒体基因组有什么特点？

20.2　人类mtDNA组成的特点是什么？

20.3　线粒体的核糖体的特点是什么？

20.4　叶绿体基因组的结构特点是什么？

20.5　叶绿体和线粒体蛋白是如何合成的？

20.6　核外遗传有何特点？

20.7　紫茉莉茎叶的白斑遗传有何特点？发生机制是什么？

20.8　核基因组通用密码子和哺乳动物及真菌线粒体的遗传密码有什么不同？

20.9　人类线粒体中的大部分mRNA没有终止密码子，而是以U或UA取而代之，解释其原因

20.10　比较一下叶绿体和线粒体蛋白的合成系统。

20.11　比较一下叶绿体和线粒体中rRNA基因的组成。

20.12　母体影响和核外遗传有何不同。

20.13　如何用实验来区分母体影响，伴性遗传和核外遗传？

20.14　酵母有几种小菌落？它们之间的区别是什么？

20.15　两个椎实螺杂交，其中一个是右旋的，结果产生的后代再进行自交，产生的F_2全部为左旋的，问亲代和F_1代的基因型应是怎样的？

20.16　果蝇有一个性连锁基因，隐性突变型称为*ma－1*（酱紫色样，maroon-like）。纯的*ma－1*雌蝇和杂合的*ma－1*雄蝇都是浅色的眼睛，这是由于缺乏黄嘌呤脱氢酶，此酶和眼色素合成有关。当杂合的雌蝇（$ma-1^{+}/ma-1^{-}$）和突变型雄蝇（$ma-1^{-}$）杂交时，所有后代的表型都是野生型。但是其中半数的雌雄后代与*ma－1*回交时产生的后代都是*ma－1*型。而另一半雌雄后代与*ma－1*回交时产生的后代都是野生型，怎样解释此结果。

20.17　一个果蝇特殊突变品系的雌蝇和野生型雄蝇杂交产生的后代中凡能成活的都是雌蝇。假设这是由于性连锁致死突变或者是一种母体遗传因子使雄性致死，那么用什么样的杂交可以加以区分？

20.18　两种品系(A,B)的果蝇进行正反交结果如下：

A♀×B♂　　后代全为雌性。

B♀×A♂　　后代全为雄性，仅少数或没有雌性。

以上现象如何解释？

20.19　有的果蝇对CO_2十分敏感，用CO_2来处理它们便可使其麻痹，敏感品系的果蝇的胞浆中有一种蛋白，它具有很多病毒的特征。抗性品系是没有颗粒的。对CO_2敏感显示了很强的母体遗传，下面两组杂交结果将如何？

(1)敏感♀×抗性♂；　(2)敏感♂×抗性♀。

20.20　有两种类型的晚樱，有下例质体效应。

O. Hookeri♀×*O. muricata*♂ ——→黄色质体

O. muricata♀×*O. hookeri*♂ ——→绿色质体

这两种类型的晚樱染色体结构相同，怎样解释以上现象。

20.21　什么叫做三系二区制种法？

20.22　植物中雄性不育的机制是什么？

第二十一章习题

21.1 什么是突变？突变有哪些类型？

21.2 在一些化学疗法中，目的是杀灭一些侵入的病原体或者癌细胞而不杀死宿主的细胞，因此必须要能区分靶细胞和宿主细胞。利用叠氮胸苷来治疗爱滋病正是应用这样的原理。请解释用叠氮胸苷来治疗时宿主细胞和HIV－1病毒有何生物学差异。

21.3 突变型 *lacZ－1* 是用吖啶处理 *E. coli* 而诱导产生的，而 *lacZ－2* 是用5BU诱导产生的它们各属于何种突变？为什么？通过对这些细胞中的半乳糖苷酶结构的研究能证实你的推断吗？

21.4 在一所医院中出生的新生儿94 075个，其中有10个人患软骨发育不全(是一种自发显性突变)。在这10个患儿中有2人其一个亲本也是患者。问软骨发育不全的突变率是多少？

21.5 自发突变是由那些因素所引起的？

21.6 放射线诱发突变的机理是什么？

21.7 化学诱变剂分为哪几类？它们怎样会引起突变的？

21.8 什么是 Ames 检测法？有何特点？

21.9 三碱基重复会引起哪些人类的遗传病？

21.10 什么叫做前突变(premutation)？是怎样形成的？

21.11 DNA 损伤修复有哪些途径？

21.12 什么是切除修复？修复的机制是怎样的？

21.13 试述 AP 核酸内切酶的修复途径。

21.14 什么是修复的 GO 系统？

21.15 试述重组修复的机制？

21.16 SOS 修复是怎样发现的？

21.17 中性突变的意义何在？

21.18 点突变对多肽结构与基因表达的影响？

21.19 突变等位基因有哪三种分类系统？

21.20 什么叫反向遗传学？

21.21 第二次突变对第一次突变会产生什么效应？

21.22 突变在进化中有何作用？

第二十二章习题

22.1 染色体结构改变是怎样产生的？

22.2 染色体畸变有几种类型？

22.3 缺失有几种类型？是怎样产生的？缺失会产生那些遗传效应。

22.4 什么是猫叫综合征？产生的原因是什么？

22.5 什么叫做拟显性？试举例说明。

22.6 重复有几种类型？是怎样产生的？有哪些遗传效应？

22.7 倒位有几种类型？倒位会产生哪些遗传效应？

22.8 易位有几种类型？怎样产生的？有何遗传效应？

22.9 什么叫做罗伯逊易位？有何遗传效应？

22.10 什么叫做位置效应？产生的机制是什么？

22.11 染色体畸变和肿瘤的发生有什么关系？

22.12 什么叫做费城染色体？它与白血病的发生有何关系。

22.13 什么叫做非整倍体？有几种类型？

22.14 单倍体和一倍体有何区别？

22.15 为什么三倍体是不育的？

22.16 人类的21－三体症是怎样产生的？

22.17 将表型为(*su*)糖质的10－3体玉米和表型为淀粉质(*Su*)的正常二倍体玉米杂交，其 F_2 代中有1758株为

Su 型，586 株为 su 型，问 *Su* 基因是否在第 10 号染色体上？

22.18　一个有色盲的男人和一个纯合正常的女人结婚后，不幸生育了两个的女儿都患有特纳氏综合征，一个视觉正常，而另一个是色盲，已知这种色盲是 X－连锁隐性性状。问这两个孩子所涉及的染色体不分离是属于父亲的 X 染色体还是母亲的 X 染色体？

第二十三章习题

23.1　重组有几种重要的形式？

23.2　断裂愈合重组模型有什么优点和不足？

23.3　模板选择模型存在哪些问题？

23.4　什么是 Holliday 模型？为什么这种模型能被广泛接受？

23.5　什么是双链断裂起始模型？有何特点？

23.6　细菌中 RecBCD 酶系在重组中起什么作用？各有哪些活性？

23.7　细菌中 RuvABC 酶系有何功能？

23.8　什么是位点特异性重组？这种重组涉及哪些蛋白？它们的功能是什么？

23.9　位点特异性重组有需要短的同源顺序，此与同源重组有何不同？

23.10　倒位重组有几种类型？

23.11　转座子是怎样发现的？

23.12　原核转座子有哪几种类型？各有什么特点？

23.13　转座子有哪几种类型？各有什么特点？

23.14　简述各种转座的机制。

23.15　以 Tn10 为例说明转座子调控的多项途径。

23.16　什么叫做多拷贝抑制？它是怎样调控 Tn*10* 的？

23.17　什么叫做顺式优先？它在 Tn*10* 转座子中起何作用？

23.18　什么叫做 *Ds* － *Ac* 因子？有何功能？

23.19　在果蝇杂种不育品系中，P 因子的作用是什么？

23.20　为什么 P × M 不育，而 P × P 和 M × P 可育？

23.21　反转录病毒整合到宿主后两端的 LTR 缺失怎样形成的？

23.22　简述反转录病毒整合模型。

23.23　什么叫做反转录转座子，试举例说明。

23.24　什么叫做内含子归巢？

23.25　I 类和 2 类内含子是如何归巢的？

23.26　免疫球蛋白的体细胞重组是怎样进行的？

第二十四章习题

24.1　什么是细胞的全能性？动物和植物细胞具有全能性吗？

24.2　克隆绵羊的意义何在？

24.3　克隆动物为什么要选择羊(而不用大鼠或小鼠)？克隆的技术关键是什么？

24.4　线虫发育有何特点？

24.5　动物的发育有几种模式？

24.6　什么是基因组等同性？转分化？愈伤组织？

24.7　什么叫胞质决定子？

24.8　什么是时序调控和异时基因？

24.9　控制果蝇的体节分化有几组基因？它们的功能是什么？

24.10　什么叫合胞体、体节和副节？

24.11　*Bicoid* 基因是怎样调节果蝇发育的？

24.12　果蝇的前后轴和背腹轴是怎样决定？

24.13 什么叫做同源异形基因?

24.14 果蝇有哪两个控制发育的复合座位?它们的结构各有什么特点?

24.15 试述 BX－C 遗传的结构与功能。

24.16 什么叫做 *Hox* 基因?其功能是什么?

24.17 简述果蝇性别决定的分子机制。

24.18 什么是初级性分化和次级性分化?

24.19 什么是 *SRY*/*sry* 基因,有什么功能?

24.20 什么叫做计量补偿作用?果蝇和哺乳动物各采用何种策略来进行剂量补偿?

24.21 简述哺乳动物性别决定机制和涉及的基因。

第二十五章习题

25.1 欧洲有一种蜗牛(*Cepaea nemoralis*)其蜗壳的颜色是复等位基因控制的,褐色(C^B)对粉红色(C^P)是显性,粉红色(C^P)对黄色(C^Y)是显性。(C^Y)对(C^P)(C^B)都是隐性的,即(C^B)>(C^P)>(C^Y)在一个群体中:褐色 236 只;粉红 231 只;黄色 33 只,共计 500 只,假设这个群体是符合哈迪－文伯格平衡群体(大的,随机交配,无选择压力),(C^B)(C^P)和(C^Y)的基因频率是多少?

25.2 在一种蛙中苹果酸脱氢酶座位上有 3 个等位基因,基因型可以用电泳来测定,基因型数如下:M^1M^1 = 8;M^1M^2 = 35;M^2M^2 = 20;M^1M^3 = 53;M^2M^3 = 76;M^3M^3 = 62 合计 254,①M^1M^2 和 M^3 在群体中的基因频率是多少?②用卡方测定来检验 MDH 的基因型是否符合哈迪－温伯格比例。

25.3 在一个大的杂交群体中 81% 的个体对于某一隐性性状是纯合的,若没有突变和选择在下一代中纯合隐性、纯合显性及杂合的个体将占多少比例?

25.4 *LetA* 对 *Leta* 基因为显性,存在的频率分别为 p 和 q,在一杂交群体中 $p+q=1$(1),若隐性表型占总数的 16%,问杂合体中的隐性基因占总的隐性基因的比例是多少?(2)若纯合隐性个体占 1%,在杂合子中隐性基因占总的隐性基因的比例是多少?

25.5 一个群体中杂合体是纯合隐性个体总数的 8 倍,隐性基因的频率是多少?

25.6 一个牛的大群体中,红色(*RR*)占 49%,杂色(*Rr*)占 42%,白色(*rr*)占 9%,(1)在此群体中亲代产生的配子中含 R 基因的占多少?(2)在另一群体中,白色的仅占 1%,9% 都是红色或杂色的,问含 r 的配子是多少?

25.7 在人类的 $S-s$ 抗原系统中由 S 和 s 共显性等位基因控制,在一个具有 3146 人的群体中基因型频率分别为 1885S,717Ss,和 224$1ss$。计算 S 和 s 的基因频率,用卡方测验来确定以上基因型频率是否符合哈迪－温伯格平衡。

25.8 在上题中在 S 座位有时还有第 3 个基因 S^u,S^u 对 S 和 s 都是隐性的,仅在纯合是才能被测出。若 S、s 和 S^u 的频率分别为 p、q 和 r,预期 S、Ss、s 和 S^uS^u 表型频率是多少?

25.9 在一个相互婚配的大的人类群体中 O 型血的人(I^i/I^i)占 60%,假定忽略突变选择,这个群体孙子辈中 O 型血占多少?

25.10 在随机交配的群体中,一个中性突变的隐性性状在男性中占 40%,在女性中占 16%,问基因的频率是多少?有多少女性是杂合的?有多少男性是杂合的?

25.11 红绿色盲是由性连锁隐性基因,约在 10000 个女性中有 64 个为色盲患者,若是随机交配群体,预期男性中色盲的比例是多少?

25.12 在群体中男性约为 8% 是红绿色盲患者(伴性隐性遗传)若在一个随机交配的群体中(1)预期女性中有多少患有色盲?(2)女性中带有色盲基因的杂合子是多少?(3)预期 2 代以后正常视觉的比例是多少?

25.13 在一个座位上有两个等位基因 Aa,U 是正向突变率为 6.0×10^{-7},V 为回复突变率为 6.0×10^{-8},若无自然选择、无迁移、无遗传漂移,当 A 和 a 突变达到平衡时频率将会怎样?

25.14 下列情况下计算有效群体的大小(Ne):(1)在繁殖种群中 50 个雄性成体和 50 个雌性成体;(2)60 个雄性成体和 40 个雌性成体;(3)10 个雄性成体和 90 个雌性成体;(4)2 个雄性成体和 98 个雌性成体。

25.15 在校园中有一个 80 只成体松鼠的群体,其 *Est* 的等位基因频率是 0.70,在附近树林里有另一群体的 Est 基因的频率为 0.5,在冬季有 20 只松鼠从树林迁移到校园里,加入了校园中的群体,在迁移之后校园里的松鼠 *Est* 基因的频率是多少?

25.16 人类中带有某种纯合的隐性常染色体基因的个体在生育前就会死去,尽管所有这种纯合体会死掉,但纯合子发生的频率逐代仍不会减少,问隐性基因固定的频率是怎样产生的?

25.17 镰刀型贫血是一种超显性,一个 *Hb－A*/*Hb－S* 的杂合体增加了对恶性疟疾的抗性,因此比 *Hb－A*/*Hb－*

A 及 *Hb - S/ Hb - S* 纯合体有更强的适应性。假设他们的适应值如下 *Hb - A/ Hb - A* = 0.88; *Hb - A/ Hb - S* = 1.00; *Hb - S/Hb - S* = 0.14。预期基因的频率是多少?

25.18　人类的软骨发育不全,是由常染色体显性基因所引起的,其突变率为 5×10^{-5},相对于正常的适应性约为 0.2。问达到平衡时此基因的频率是多少?

25.19　在一个特定人类群体中随机地采集了 100 个人的 DNA,用限制性酶 *Bam*H Ⅰ处理,经电泳后进行 Southern 杂交,有的 DNA 样品(56 人)出现了单条 6.3kb 的带,另一些(6 人)显示单条 4.1kb 的带,还有一部分(38 人)显示了 6.3kb 和 4.1kb 两条带。(1)以 *Bam*H Ⅰ位点来解释此结果;(2)限制性位点的等位频率是多少?(3)这个群体的限制性位点是否符合哈迪 - 温伯格平衡?

25.20　在一个池塘里有 50 只虎螺,用这些虎螺为材料进行蛋白电泳来研究遗传变异。每个虎螺选择 5 个表型座位来分析,那就是 AmPep、ADH、PGM、MDH 和 LDH - 1。结果发现 AmPep、ADH 和 LDH - 1 这 3 个位点上未发生变异,而在 PGM 和 MDH 两个座位上有变异,具体情况如下:

基因型	个体数	基因型	个体数
AA	11	*DD*	35
AB	35	*DE*	10
BB	4	*EE*	5

计算多态位点的比率和群体中杂合子的比。

习题参考答案

第一章

1.1 (1)红;(2)3红,1黄;(3)全红;(4)1/2红,1/2黄。

1.4 F_2 基因型比(C是有色,c是无色)是1/4CC∶1/2Cc∶1/4cc。若我们考虑到有颜色的植物,则CC纯合子=1∶2。若有色植物是随机包装的,那么CC的概率为1/3(即有色植物中1/3为纯合)而Cc的概率为2/3。

1.5 (1)亲代是 *Rr*(粗糙)和 *rr*(光滑);F_1 是 *Rr*(粗糙)和 *rr*(光滑)(2)$Rr \times Rr \rightarrow$3/4粗糙,1/4光滑。

1.7 后代比近似3∶1,故亲代是杂合的。在显性性状的后代中纯合∶杂合=1∶2,所以其中有1/3是纯种。

1.8 黑色对棕色是显性。若 *B* 是黑色等位基因,*b* 是棕色等位基因,那么雌性X是 *Bb*,雌性Y是 *BB*,雄性是 *bb*。

1.10 (1)黑色 F_1 后代的互交的形式有4种可能 $BB \times BB$, $Bb \times BB$, $BB \times Bb$, $Bb \times Bb$,而产生白色后代的仅最后一种组合。黑色杂合体的概率为 $2/3(2Bb)/(BB+2Bb)$。所以产生白色个体的概率为 $2/3 \times 2/3 \times 1/4 = 1/9$。(2) $2/3 \times 1 \times 1/4 = 1/6$;(3)最大的可能 F_1 bb雄性和PBb雌性回交。BB×bb→50%的后代为Bb,另外50%的后代为白色bb。

1.13 (1) *WWDd* × *wwdd*; (2) *Wwdd* × *Wwdd*; (3) *wwDD* × *WWdd*; (4) *WwDd* × *wwdd*; (5)WwDd × Wwdd

1.14 杂交组合为 *AaBbCc* × *AaBbCc*(1)产生ABC表型的概率为 $(3/4)^3 = 27/64$;(2)产生 *AABBCC* 基因型的概率为 $(1/4)^3 = 1/64$

1.17 (1)F_1 的基因型为 $AaBbCC^s$,表型都是野鼠色,黑色。F_2 有27/64野鼠色,黑色;9/64野鼠色,黑色,喜马拉雅白化;9/64黑鼠色,棕色;*3/64* 黑鼠色,棕色,喜马拉雅白化;3/64黑色,喜马拉雅白化;3/64棕色;1/64棕色,喜马拉雅白化。(2)27/64的F2是野鼠色,黑色。$AaBbCC^s \times AaBbCC^s$ 产生出的 F_2 中,基因型为 $AaBBCC^s = 1/2 \times 1/4 \times 1/2 = 1/16 = 4/64$。所以F2中鼠色,黑色的比例是 $AaBBCC^s = 4/27$。(3)F_2 中喜马拉雅白化=9/64(野鼠色,黑色,喜马拉雅白化)+3/64(黑鼠色,棕色,喜马拉雅白化)+3/64(黑色,喜马拉雅白化)+1/64(棕色,喜马拉雅白化)=16/64。所以有3/64+1/64=4/64的棕色,喜马拉雅白化。因此在 F_2 代中的喜马拉雅白化有棕色斑的为1/4。(4) F_2 鼠色=27/64(鼠色,黑色)+9/64(野鼠色,黑色,喜马拉雅白化)+9/64(黑鼠色,棕色)+3/64(黑鼠色,棕色,喜马拉雅白化)=48/64。F_2 代中有27/64+9/64=36/64为黑色,鼠色。因此 F_2 鼠色中有黑斑的比例为3/4。

第二章

2.1 (3)

2.3 (3)

2.5 (1)是的。若在一物种存在有性交配,两个单倍体细胞能融合产生一个二倍体细胞,这个二倍体细胞经减数分裂产生单倍体后代。例如脉孢菌和酵母都是如此。(2)不。因为单倍体细胞不能进行减数分裂,或只能产生一种假的减数分裂,如雄蜂。

2.7 (3)

2.9 (1)中期;(2)后期

2.13 (1)$1/2 \times 1/2 \times 1/2 = 1/8(2)1 - 1/8 - 1/8 = 6/8 = 3/4$

2.14 1条长染色体和短染色体可能是异形配对染色体,如X和Y染色体。

2.15 (1)不对。因表皮细胞是体细胞,是不能进行减数分裂的。(2)不会。因为形成配子的染色体是独立分配,自由组合和有交换发生的,而一对同源染色体所携带的各等位基因不会都相同,所以产生的配子完全相同的概率极小。

第三章

3.1 $(1/2)^5 = 1/32$

3.2

	外祖母	外祖父	祖母	祖父
X染色体	+	+	—	—
Y染色体	—	—	+	+

3.4 所有女儿都正常,儿子中将有一半是色盲,一半为正常。3.6 所有的子女都正常。(C^+C^+雌性和 C^+Y 雄性),(两性皆为 a^+a)。

3.7 亲代为 ww vg^+ vg^+(雌)和 w^+y vg vg(雄)。(1)F_1 雄性皆为 wy vg^+ vg白眼长翅;F_1 雌性全部为 w^+w vg^+ vg

红眼长翅。(2)F_2 雄性 3/8 红眼长翅;3/8 白眼长翅;1/8 红眼残翅;1/8 白眼残翅。雌性的各种表型比例相同。(3)F_1 雌性与 P_1 回交:$wy\ vg^+\ vg \times ww\ vg^+\ vg^+ \rightarrow$所有后代都是白眼长翅。$F_1$ 雄性与 P_2 回交:$w^+w\ vg^+\ vg \times w^+y\ vg\ vg \rightarrow$雄性后代,全部为红眼,一半长翅,一半残翅,雄性后代:1/4 红眼长翅;1/4 红眼残翅;1/4 白眼长翅;1/4 白眼残翅。

3.10　最简单的假设是棕色牙是由性连锁显性突变等位基因。A 个体的基因型为 *BY*,其妻为 *bb*,所有的儿子为 *bY*,表型正常,不表现为棕色牙。女儿从父亲那儿接受了带有突变基因 *B* 的 X 染色体,因此她们的基因型为 *Bb*,表型为棕色牙。女儿的儿子一半接受了带有突变基因 *B* 的 X 染色体为棕色牙,一半接受了带有正常基因 *b* 的 X 染色体为白色牙。

3.12　(1)生了一个患囊性纤维化的孩子,双亲必定是囊性纤维化这个常染色体隐性突变基因的异合体,所以下一个孩子患此病的机会为 1/4;(2)生一个携带者(表型正常,带有有害基因)的机会为 2/3。

3.18　(1)可以排除 X - 连锁隐性基因,因这种遗传的女性患者极少。如 I - 2 为患者,则此性状应传给所有的儿子,与图谱不符。II - 1,II - 2 结婚应是所有的子女都获得此性状,也与图不符。常染色体隐性也可被排除,因 II - 1 和 II - 2 结婚按常染色体隐性遗传规律,他们的全部子女都应为患者。(2)此图谱可能是 X - 连锁显性或者常染色体显性,常染色体显性的特点:(1)患者的亲代总有一方为患者;(2)几乎每代都有患者;(3)患病与否和性别无关;(4)亲代正常后代全部正常;(5)双亲为患者,后代约有 3/4 为患者。除第 4 条外,图谱满足了以上条件,故可能是常染色体显性。X - 连锁显性的特点是:(1)女性患者的儿子和女儿都有一半为患者,III - 2 婚配符合此条;(2)男性患者的女儿应该全部为患者,儿子正常,II - 6,II - 7 婚配符合此条;(3)若父母皆为患者,女儿全部为患者,儿子半数为患者,II - 1,II - 2 婚配符合此条。

第四章

4.2　可能的基因型是:*w/w*, *w/w1*, *w/w2*, *w1/w1*, *w1/w2*, *w2/w2*

4.8　后代中一半为杂色(R/r),1/4 为红色(R/R),1/4 为白色(r/r)

4.9　(1)*F/Fo/oXf/fO/O*,F_1 为 *F/fO/o*,为表皮有绒毛,圆坚果。(2)F_1 自交,F_2 中有 3/16 有绒毛,卵形坚果;6/16 有绒毛,圆坚果;3/16 有绒毛,无坚果;1/16 表皮光滑,卵形坚果;2/16 光滑,圆坚果;1/16 光滑,无坚果。(3)*F/fO/o* × *f/fO/o* 杂交,产生的后代 1/4 有绒毛,卵形坚果;1/4 有绒毛,圆坚果;1/4 表皮光滑,卵形坚果;1/4 光滑,圆坚果。

4.15　在后代中,显示不分离,选择植物必须是纯合的。基因型包括 9/16 有色植物:
1A/AB/B:*2A/aB/B*:*2A/AB/b*:*4A/aB/b*,在这些基因型中只有一种是纯合的,所以答案是 1/9。

4.17　(1)若 *Y* 代表黄色,*y* 代表鼠色,而 *Y/Y* 致死,*Y/y* 为黄色,*y/y* 为鼠色。*C* 决定有色素,决定无色素(白色),亲代的基因型为 *Y/yC/c*(黄色)和 *Y/yc/c*(白色)。(2)比例是 2 黄:1 鼠色:1 白色。没有一只黄色的是纯种。因为纯合个体不能存活,它们都是杂合的。

4.18　(1)*Y/YR/R*(深红) × *y/yr/r*(白色)→F_1 代植物为紫玫瑰红色(*Y/yR/r*)。F_1 自交产生的 F_2 代 1/16 深红(*Y/YR/R*),2/16 桔红(*Y/YR/r*),1/16 黄色(*Y/Yr/r*),2/16 紫红(*Y/yR/R*),4/16 紫玫瑰红(*Y/yR/r*),2/16 淡黄(*Y/yr/r*),4/16 白色(*y/yR/R*, *y/yR/r* 和 *y/yR/r*)。F_1 杂种与深红亲本回交产生 1/4 深红,1/4 桔红,1/4 紫红,1/4 紫玫瑰红。(2)*Y/YR/r* × *Y/yr/r*→1/4 桔红(*Y/YR/r*),1/4 黄色(*Y/Yr/r*),1/4 紫玫瑰红(*Y/yR/r*),1/4 淡黄(*Y/yr/r*)。(3)*Y/Yr/r*(黄色) × *y/yR/r*(白色)→1/2 紫玫瑰红(*Y/yR/r*),和 1/2 淡黄(*Y/yr/r*)。

第五章

5.2　经 χ^2 测验,$\chi^2 = 16.10$;$p < 0.01$;$N = 3$,结果表明两个基因不符合 1 : 1 : 1 : 1 的比例,可能由于有连锁的存在。若考虑每一座位的分离情况,*B/- : b/b* 的比例为 203 : 197,符合 1 : 1;但 *A/- : a/a* 之比为 240 : 160,不符合 1 : 1。

5.5　子代 158 只小兔中,重组类型是英国长毛兔和非英国短毛兔。图距为
$(11 + 6)/158 \times 100\% = 10.8\% = 10.8$m.u.。

5.7　幸运的是在染色体上存在一个双突变,这样才能排除 *vg* 和 *m* 之间的连锁。因为在雄性果蝇中不发生交换,所以在 F_1 雄性中无法产生 vgm 配子,预计:

P　　$vg^+/vg^+\ m/m \times vg/vgm^+/m^+$

F_1　　$vg^+/vgm/m^+$(野生型)

这种杂交的结果表明 m 不是 X - 连锁的。F_1 的杂交是双因子杂种的杂交,我们可以预期 F_2 中:$vg^+/m^+ : vg/m^+ : vg^+/m : vg/m = 9 : 3 : 3 : 1$。但在这样小的样本中可能会有变化。这表明 *vg* 和 *m* 是独立分配的。若 *vg* 和

m 连锁，那么情况如下：

P　　$vg^+m/vg^+m \times vg^+m/vgm^+$

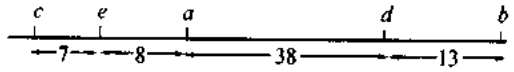
F_1　　vg^+m/vgm^+（野生型）

$F_1 \times F_1$ 结果如下表：

	vg^+m(雄性)	vg/m(雄性)
vg^+/m(雌性)	深紫色	野生型
vg/m^+(雌性)	野生型	残翅
vg^+/m^+(雌性)	野生型	野生型
vg/m(雌性)	深紫色	残翅

表型比例为野生型：残翅：深紫：残翅深紫 = 4：2：2：0。在这样的杂交中我们无法计算出两个突变位点的距离。因此如果两个基因连锁，在 F2 中就得不到双突变的后代。M 不在第 2 号染色体上。

5.8



5.9　45%ab⁺，45%a⁺b，5%a⁺b⁺，5%ab。

5.10　(1)0.035 + 0.465 = 0.50；(2)所有子代表型为 a^+b^+。

5.13　(1)F_1 代是 AB/ab。F_1 代产生的配子是 40%*AB*；40%*ab*；10%*Ab*；10%*aB*。从测交 $F_1 \times ab/ab$ 得到 40%*AB/ab*；40%*ab/ab*；10%*Ab/ab*；10%*aB/ab*。(2)F_1 代是 Ab/aB。F_1 代产生的配子是 40%Ab；40%aB；10%ab；10%AB。从测交 $F_1 \times ab/ab$ 得到 40%*Ab/ab*；40%*aB/ab*；10%*ab/ab*；10%*AB/ab*。

5.15　每对染色体是独立分配的。我们从已知图距中可以计算出配子的相对比例(P:亲组合；R:重组合)。

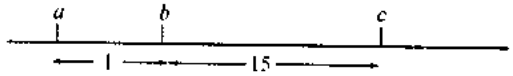
P	*R*	*P*	R	*P*	R
*AB*0.4	*Ab*0.1	*CD*0.45	*Cd*0.05	*EF*0.35	*Ef*0.15
*ab*0.4	*aB*0.1	*cd*0.45	*cD*0.05	*Ef*0.35	*eF*0.15

F1 中产生各种配子的概率：

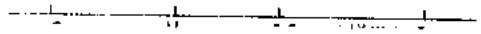
(1)*ABCDEF* = 0.4 × 0.45 × 0.35 = 0.063(6.3%)　(2)*ABCDef* = 0.4 × 0.05 × 0.35 = 0.007(0.7%)

(3)*AbcDeF* = 0.1 × 0.05 × 0.15 = 0.00075(0.075%)　(4)*aBCdef* = 0.1 × 0.05 × 0.35 = 0.00175 (0.175%)(5)*abcDeF* = 0.4 × 0.05 × 0.15 = 0.003(0.3%)

5.19

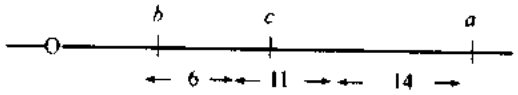


5.21　*arg*　　*met*　　*thi*



5.22　这两个基因可被认为是不连锁的，因为亲型(20)近似等于非亲型(18)数目。

5.25　(1)应用四分子分析公式，图距 = (1/2 · T × NPD)/总数 × 100%，a - b 距离是 19.6mu；b - c 为 11mu；a - c 为 14mu。(2)着丝粒距离从下列公式可计算：(第二次分裂分离四分子%)/2



第六章

6.1　(1)头宽平均值 = 3.15，标准差为 0.49；翅长平均值 = 35.21，标准差为 7.68。(2)相关系数 γ = 0.973(3)头宽和翅长有着强的正相关。

6.2　每个纯种亲代的基因是纯合的，表型应相同，如产生差异则是环境引起的；F_1 代是杂合的，但基因型相同，表型也应相同，如产生差异也是由环境造成的，所以 F_1 代之间的差异，不比每一亲本内的差异大。

6.4　(1)8cm；(2)1(2cm)：6(4cm)：15(6cm)：20(8cm)：15(10cm)：6(12cm)：1(14cm)。(3)F_2 中 *AABBCC* 为 $(1/4)^3$；*aabbcc* 为 $(1/4)^3$，和亲代的高度相等，总的比例为 $(1/4)^3 + (1/4)^3 = 2/64$。(4)预期 20/64 的 F_2 高度为 8cm。

6.6　是数量性状。因为变异是遗传的，F_1 的平均值在双亲平均值之间。F_2 表现出连续变异，变异范围在两亲本表型之间的区域。

6.9　$(1/4)^{20}$

6.10　选择差异是 0.14g；选择反应是 0.18g，$h^2 = 0.08/0.14 = 0.57$

6.11　遗传力为 0。因在群体中遗传变异性不影响血压。

6.12　选择差异 = 14.3 - 9.7 = 4.6；选择反应 = 13 - 9.7 = 3.3；狭义遗传力 = 2(b) = 2(0.46) = 0.92。

6.13　(1)父本平均值 = 71.9 方差 = 11.6；儿子平均值 = 70 方差 = 10.2；(2)相关系数 = 0.49(3)斜率 = b = 0.46

6.14　选择反应 = 狭义遗传力 X 选择差异

选择反应 = 0.60 × 10g = 0.92。

6.15　(1)$F = 2(1/2)^6 = 1/32$；　$F = 2(1/2)^5 + (1/2)^3 = 3/16$

第七章

7.1　(1)主要用于确定亲代是否是携带者及产前诊断(2)由于过氧化物酶缺陷的患者，携带者和正常人的酶活性区域的明显不同，所以对于个案病人，只要准确地测出其酶活性，就可以说明被测个体是否正常。而半乳糖酶则不然，携带者和正常人的酶活性值部分重叠，如一个人的酶活在 25 ~ 30 单位，就很难判断是正常人还是携带者。

7.2　一种是胚系突变，这种可能性比较小。较大的可能性是双亲的突变位点不同，可以产生互补。

7.3　他们的孩子是正常的，*Aabb* × *aaBB*→*AaBb*

7.7　(1)*c* - *e* - *d* - *a* - *b*(2)表明缺失 4 跨越了两个顺反子，这不影响(1)的结论。

7.9　酶基因的多效性不能否定一基因一酶。当一个基因发生突变影响到其产物的功能，而这个产物可能是酶、结构蛋白，运载蛋白，受体蛋白等。它们本身发生了改变又会影响到一系列其他代谢或生理功能，出现多效性。多效性的有些表型和基因是间接的关系，不能说明一个基因控制多个蛋白。

7.10　(1)60%的减数分裂未发生交换(2)20mu；(3)10%；(4)→B→A→Val

↑　↑

1　2

7.11　(1)1 × 2→F_1 全部紫色，表明两个基因涉及同一个生化途径，并可以互补。设 1 为 *aa*，2 为 *bb*。则

aaBB × *Aabb*→*AaBb*

白　白　紫色

第一种情况：1 × 2→F_1 全部紫色→F_2：1/2 紫色，1/2 白色

图解：P　*aB/aB* × *Ab/Ab*→F_1：*aB/Ab*(全紫)→F_2：1*aB/aB*(白)：2*aB/Ab*(紫)：1*Ab/Ab*(白)

第二种情况：1 × 3→F_1 全部紫色→F_2：9(紫)：7(白)

P：*aB/abD/D* × *aB/aBdd*→F_1：*aB/ABD/d*(紫)→F_2：9(*A*_ *BBD*)紫：3(*aaBBD*_)白：

3(*A*_ *BBdd*)白：1(*aaBBdd*)白

第三种情况：1 × 4→F_2：白→F_2：白

P：*aB/aBD/D* × *aB/aBD/D*→*aB/aBD/D*(白)→F_2 同 F_1

(2)2 × 3→*Ab/AbD/D* × *AB/Abdd*→F1：*Ab/ABD/d*(紫色)

F_2：9(*A*_ *B*_ *D*_)紫：3(*Ab/AbD*_)白：3(*A*_ *Bdd*)白：1(*Ab/Abdd*)白

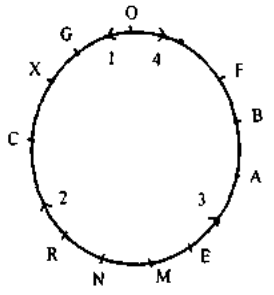
2 × 4→同 1 × 2

第八章

8.1　A 品系是 *thrleu*⁺，B 品系是 *thr*⁺ *leu*⁻，若 B 品系被 A 品系转化应出现 *thr*⁺ *leu*⁺，因此可以用基本培养基加以检测筛选。

8.3　因此 $F^+ \times F^-$ 的杂交中完整的 *F* 因子经滚环复制进入受体，使受体也获得了 *F* 因子，故以很高的频率变成供体。在 *Hfr* × F^- 中的 *Hfr* 品系和 F^+ 不同，其 *F* 因子也整合在宿主的染色体上。DNA 转移到受体时，是以 *F* 因子的 OriT 为前导，最末端是 F 因子的致育基因，在 DNA 转移过程中由于接合管的随机断裂常使致育基因不能转移到受体细胞中，所以受体仍为 F^-，很少能成为 F + (供体)。

8.4



8.8 (1)GT;(2)ST;(3)ST;(4)GT;(5)GT;(6)ST;(7)ST;(8)B;(9)N

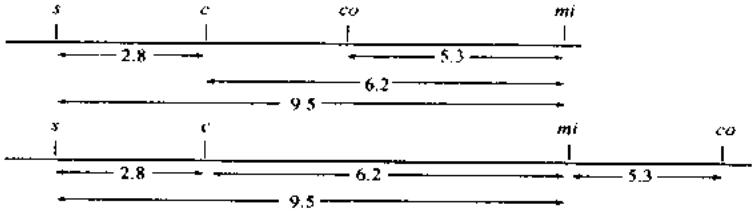
8.9 因 *aceF*$^+$ 和 *dhI* 的共转导频率较高，所以它们是紧密连锁的。

8.10 *trpB* 标记与 *cys*$^+$ 选择标记共转录频率高于 *trpE* 标记，因此，*trpB* 和 *cysB* 紧密连锁。

8.11 顺序为 *supD* – *cheA* – *cheB* – *eda*。

8.12 因 0.1ml 的噬菌体可产生 20 个噬菌体，即 200 个噬菌体/ml，稀释的浓度是 10^{-8}，所以原来的浓度为 200×10^8 或 2×10^{10}。

8.13 根据资料可能有两种情况：



8.14 重组数为 0.5%(噬菌体数是很小的，这个值是粗略地估计)。

8.15 A、D 和 F 在一个基因中，B 和 G 在第二个基因中，C 和 E 在第三个基因中。

第九章

9.2 (4)

9.4 5′端的 C 原子和磷酸键相连。

9.7 从各种生物中提取的 DNA 双链中，只有 4 种碱基组成两种碱基对。在各种情况下，A 的量总是和 T 的量相等；G 的量和 C 的量相等。不同的 DNA 尽管 G – C 的含量不同，但总是 A = T，G = C，但 (A + T) ≠ (G + C) 以上表明 A – T 和 G – C 形成碱基对，另外从几何形状和氢键的长度来看也是能是 A – T 和 G – C 配对。

9.12 这种噬菌体是单链 DNA。

9.14 因 DNA 上每个位点上存在某一碱基的概率为 1/4，选用 4 个碱基组成的酶切位点在 DNA 可能每隔 256bp 出现一次，因此应选识别 4 个碱基的限制酶，可将 DNA 切成约 250bp 的短片段。同理若想切成 4kb 的随机片段可选识别 6 个碱基的限制性酶[$(1/4)^6 = 1/4096$]

9.16 GUUA 的概率是 $0.3\times0.25\times0.25\times0.20 = 0.00375$

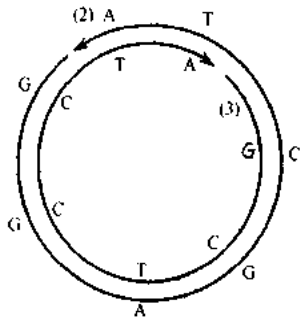
$10^4\times0.00375 = 3750$. 所以 10^4nt 长的 RNA 可能出现 GUUA 约为 3750 次

第十章

10.22 (1)如图所示：

2)GCTCCTA

CGAGGAT



10.23 (1)T4 染色体是环形的，末端是冗余的。在变性和复性时，很少有完整的分子，但大部分分子会出现单链末端，带有很长单链末端的分子可能形成复杂的环形结构。(2)T7 染色体是单一顺序，所有的分子都相同，因此所有的分子都是完整的。(3)λ 染色体是线状的，每条染色体的顺序都相同，末端有 COS 呈粘性末端状。因此所有的分子都有单链末端或呈环状。

10.24 Barr 小体：兼性异染色质；着丝粒：组成型异染色质；端粒：组成型异染色质；大部分表达基因：常染色质。

第 11 章

11.1 *E.coli* 的 OriC 长 245bp。(1)富含 A – T；(2)含有多个回文 GATC；(3)有 4 个 9bp 的反向重复顺序，3 个 13bp

正重复顺序，

11.2　DnaA：OriC 的识别和解链；DnaB 和 DnaC：解旋酶，解开双链；

TopII：复制叉的移动；SSB：使单链稳定；DnaG：合成引物。

11.4　δ，δ′，φ，χ，γ－五聚体，是 β 亚基的装配器。β－2×β2（两个二聚体），分别夹住前导链和后滞链。2×（α，ε，θ）－核心酶，聚合功能。2×J 亚基－是将两（α，ε，θ）三聚体连接起来，对称排列，分别合成前导链和后滞链。

11.16　关键$^{15}N-^{15}NDNA=HH$；$^{15}N-^{14}NDNA=HL$；$^{14}N-^{14}NDNA=LL$。（1）第一代：全部为 HL；第二代：1/2HL，1/2LL；第三代 1/4HL，3/4LL；第 4 代 1/8HL，7/8LL，第 6 代 1/32HL，31/32LL；第 8 代 1/128HL，127/128LL。（2）若是全保留：第一代 1/2HH，1/2LL，第 2 代 1/4HH，3/4LL；第 3 代 1/8HH，7/8LL；第 4 代 1/16HH，15/16LL；第 6 代 1/64HH；63/64LL；第 8 代 1/256HH，255/256LL。

第十二章

12.1　（1）转录仅一条模板链；（2）RNA 杂合双链不稳定；（3）转录不需要引物；（4）底物不同；（5）酶系不同；（6）转录的开链区不扩大，只移动；（7）RNA 产物会从模板上释放出来。

12.2　RNA pol 有 2 个 α 亚基、β 和 β′亚基组成核心酶，核心酶加 σ 亚基形成全酶。功能（1）直接识别启动子。（2）使 DNA 解链；（3）合成 RNA；（4）确定转录方向；（5）识别终点；（6）不能校正。

12.3　（1）结构典型由－35、－10 和转录起始位点构成；（2）顺序保守；（3）位置和距离恒定，（4）直接和 RNA pol 结合；（5）决定转录的方向和强弱；（6）有回文顺序；（7）常和操纵子相邻，受到调控；（8）仅位于基因的 5′近上游；（9）顺式作用。

12.5　RNA pol 核心酶在 DNA 上像尺蠖一样爬行，即（1）覆盖 35bp，延伸时前部不动，后部逐步收缩，向前移动了 9 个 bp，同时 RNA 从 RNA pol 中露出 9 个 nt。（2）后部不动，前部一次向前延伸 9bp，使核心酶又重新覆盖 35bp。（3）第二轮收缩，伸长重新开始。

12.6　有两种：（1）强终止子又叫内部终止子；（2）弱终止子又叫 ρ 依赖性终止子；强终止子的特点是：（1）有回文，可形成茎环结构，（2）富含 G－C，（3）3′端有寡聚 U 的存在。弱终止子，长 50～90nt，在终止子的上游，特点是（1）富含 C，缺乏 G，（2）终止效率随富 C/贫 G 的长度增加而增加。

12.8　Ⅱ类启动子由以下的元件构成

Ⅱ类启动子元件	核心元件	*TATA* 框
		InR（起始子）－Py_2CAPy_5（－3～＋5）
	上游元件（UPE）	共同的 UPE：*CAAT* 框、*GC* 框、*Oct*、*κB*、*ATF* 可诱导的 UPE：HSE、GRE、MRE、TRE、SRE
	远上游元件	增强子、减弱子、静息子、UASs

第十三章

13.9　真核生物 mRNA 的内含子有Ⅰ类，Ⅱ类，核 mRNA 以及反式剪接四类，其特点如下表所示。

	Ⅰ类内含子	Ⅱ类内含子	核 mRNA 内含子	反式剪接内含子
边界顺序	5′U↓-G↓3′	↓GU-AG↓	↓GU-AG↓	-↓GU-，-A-AG↓
特殊顺序	中部核心序列（PQ）（RS），IGS	分枝顺序 保守 A	分枝顺序 保守 A	保守 A
参与剪接的因子	GTP、镁离子、核酶	镁离子、核酶	U_1，U_2，U_5，U_6	U_2，U_4，U_6
中间型	环状	套索	套索	Y 型分子

13.10　主要有两种策略，一是通过形成茎环结构，如Ⅰ，Ⅱ型内含子，二是通过 RNP 形成剪接体，如核 mRNA 和 SLRNA 反式剪接。

13.12　转酯反应是指酯键从某一位转移到另一位置的化学反应的过程。在Ⅰ类，Ⅱ类内含子的核 mRNA 和 SLRNA 反式剪接以及编辑都属于转酯反应。

第十四章

14.6　tRNA 的二级结构为三叶草状，其特点是：（1）有 22 个碱基恒定；（2）5′端配对；（3）3′－OH 永远是单链区；（4）有三个茎环（DHU 环，tψC 环，反密码子环）和氨基酸臂；（5）额外环；（6）GU 配对。

14.7 tRNA 的三维结构为 L 型，特点是(1)受体臂位于 L 型的一侧；(2)TψC 环和 DHU 环形成 L 型的转角；(3)在一些保守半保守的碱基之间形成很多的三级氢键(在 2′-OH 上)；(4)碱基的朝向产生堆积；(5)在反密码子环中很少有三级氢键，反密码子区的朝向可能发生改变。

14.10 一些突变型 RNA 能够恢复编码蛋白的基因的突变效应，即 tRNA 的反密码子发生突变可以识别某些无义突变的密码子，在无义突变位点插入一个氨基酸，使蛋白质的合成得以恢复而不致截短，此突变的 tRNA 基因称为抑制基因或校正基因。抑制基因可以分为无义抑制和错义抑制两类，能抑制无义突变的称无义抑制，而抑制错义突变的则称为错义突变。

14.11 抑制基因一般不会改变正常的终止。这是由于(1)抑制基因识别的终止密码子不一定和正常终止密码子相同；(2)释放因子和抑制基因竞争和终止密码子的结合；(3)抑制基因的效率很低，通常为 1% ~ 5%，故常不会抑制正常的终止。

14.12 细菌核糖体的活性位点其功能和位置如下表：

活性位点	功能	位置
mRNA 结合位点	结合 mRNA 和 IF 因子	30S 亚基 P 位点附近
P 位点	结合 fMet-tRNA 和肽酰 tRNA	大部分在 50S 亚基
A 位点	结合氨基酰-tRNA	大部分在 30S 亚基
E 位点	结合肽酰 tRNA	50S 亚基
5S RNA	和 32S rRNA 结合	P 和 A 位点附近
肽酰转移酶	将肽链转移到氨基酰-tRNA 上	50S 的中心突起
EF-Tu 结合位点	氨基酰-tRNA 的进入	30S 外部
EF-G 结合位点	移位	50S 亚基的界面上，L7/L12 附近

第十五章

15.1 氨基酸进入蛋白质合成途径是通过氨基酰 tRNA 合成酶．这种酶将氨基酸和特异的 tRNA 连接起来。所有这种合成酶的功能都是分步进行的，其机制如下：

活化：aa + ATP + E→aa - AMP - E + PPi

转移：aa - AMP - E + tRNA→aa - tRNA + AMP

15.3 三位点学说认为在细菌的核糖体中除了 A、P 两个位点以外还有 "E" 位点，蛋白质合成发生移位后，空载 tRNA 并不直接从核糖体的 P 位点排出，而是进入 E 位点。E 位点和 A 位点是反协同效应的，即当 A 位点被 aa-tRNA 占据时，E 位点的亲和力降低，而 E 位点被空载 tRNA 占据时，A 位点的亲和力降低。

15.5 (1)识别起始密码子 AUG，在细菌有时为 GUG，UUG，(2)和延伸 $tRNA^{Met}$ 结构不同。(3)携带的氨基酸是甲酰化的，阻断氨基，使其不能参与延伸。(4)进入 P 位点，而不是 A 位点。

第二十章

20.15 亲代和 *F1* 代的基因型应为 ♀ $Dd \times d \rightarrow dd$

20.17 第一种假设性连锁致死基因 L 是显性的，在雌性中纯合，在雄性中致死，在雌性中不致死，可以设计以下杂交：(P)♀性 $LL \times + y$ ♂ $\rightarrow L+ : Ly$(死)　(F_1)

$$L+ \times + y \rightarrow L+ : Ly\text{(死)} : + y \quad (F_2)$$

结果♀ : ♂ = 2 : 1

第二种假设是有卵的细胞质所传递的遗传，在这种情况下，上述杂交 F_2 代中也只有雌蝇可以成活，雄性全部死亡。

第二十一章

21.3 吖啶是一种插入剂，所以它引起的突变是移码突变，而 5BU 是 T 的替代物，所以引起的突变是点突变。通过正常的 β-半乳糖苷酶的一级结构和突变体半乳糖苷酶的比较，可以确定以上的推测。

21.19 分为三类：(1)功能的获得与丢失，适用于单倍体与二倍体生物，含功能丢失等位基因和功能获得的等位基因；(2)显性关系，只适用于二倍体，等位基因分为显性，隐性，部分显性，共显性。(3)Muller 分类：只适用于二倍体，分为无效(无活性)，亚效(渗透突变)，反效，超效，新效等位基因。

第二十二章

22.17　*Su* 基因不在第 10 条染色体上，因后代分离比为正常的 3 : 1，若在第 10 条上将不可能得到 3 : 1 的分离比。

22.18　带有色盲的特纳氏综合征的女孩是由于母亲的 X 染色体不分离，产生两种配子，一种是 XX，一种是"O"，后者与来自父亲带有色盲基因的 X^c 结合形成 X^cO。而有正常视觉的特纳氏综合征的女孩是由于父亲带有色盲的 X 染色体和 Y 染色体同趋向一极，母亲正常的 X 染色体和不含 X 染色体的精子结合产生了 XO。

参 考 文 献

总的参考文献

杜传书,刘祖洞主编.1992.医学遗传学(第二版).北京:人民卫生出版社

刘祖洞.1991.遗传学第二版(上、下).北京:高等教育出版社

盛祖嘉,沈仁权.1988.分子遗传学.上海:复旦大学出版社

Brown,T.A.(1992), *Genetics a Molecular Approach*, 2 nd ed. Chapman and Hall,London

Griffiths,A.J.F.,Miller,J.H.,Suzuki,D.T.,Lewontin,R.C.,Gelbart,W.M.(1993), *An Introduction to Genetic Analisis*, 5 th ed.W.H.Freeman and company /New York

Lewin, B. (1997), *Genes* Ⅵ .Oxford University Press, Oxford

Maxon, L.R. and Daughherty, C.H. 1992. *Genetics a Human Perspective* 3rd ed ,William Brown

Primrose,S.B.(1955), *Principles of genom Analysis*. Blackweii, Oxford

Ridley, M.and Berg,P.(1991), *Genes and Genome*, University Science Books, Mill Velley,CA,USA

Russell,P.J.(1992),*Genetics*,3rd ed.Harper Collins Publishers

Strachan, R.F. and Read, A.P. (1996), *Humen molecular Genetics* , BIOS Scientific Publishers, Oxford

Winter, G.I. 遗传学(影印版).科学出版社,1999

第一章 孟德尔定律

G.孟德尔(1866).植物杂交实验.载:G.孟德尔等著(梁宏,王斌译).1984.遗传学经典论文集.北京:科学出版社

I.夏恩,S.朗比尔著(李振华等译).1987.遗传学先驱摩尔根.北京:科学出版社

Grant,V. (1975), *Genetics of Flowering Plants*. New York: Columbia University Prees

McKusick, V.A. *et al*.(1990), *Mendlian Inheritance in Men*, 9 th ed.Baltimoreand London: Johns

Hopkins Press.The "Bible"of medical genetics: a 2000 - Page compendium of all the known humen inherited disorders, both autosomal and X - linked

Mange, A.P. and E.J.Mange.(1990) *Genetics Humen Aspect*,2 rd ed..Sunderland , Mass Sinauer. A Useful and up - to - date text for a course in humen genetics

第二章 遗传的染色体学说

Sutton,W.S.(1903).染色体遗传.载:G.孟德尔等著(梁宏,王斌译).1987.遗传学经典论文集.P37.北京:科学出版社

McLeish, J. and B. Snoad. (1958), *Looking at Chromosomes*. New York: Macmillan. A short classic book consisting of many superb photos of mitosis and meiosis

Crow, J.F.(1983), *Genetics Notes*. 8 th ed. New York: Macmillan. A short and concise review of genetics from Mendel to populations

Von Wettstein, D., *et al*. 1984. *The Synaptonemal Complex in Genetic Segregation*. Annual Review of Genetics 18: 331 - 414. An up - to - date review of the structure and function of the complex

McIntosh,J.R.,and McDonld.K.L.(1988). *The Mitotic Spindle*. Scientific American (October)

第三章 性别决定和性相关遗传

Bridges, C.B.(1925).性别对染色体和基因的关系.1987.载:G.孟德尔等著(梁宏,王斌译)遗传学经典论文集,P67.北京:科学出版社

Wilson,E.B.(1905).昆虫染色体与性决定的关系.1987.载:G.孟德尔等著(梁宏,王斌译)遗传学经典论文集,P37.北京:科学出版社

McClung, C. E. 1902. *The accessory chromosome - sex determinant?* Biol. Bull. 3: 43 - 84

Wilson, E. B.(1905). *The chromosomes in relation to the determination of sex in insects*. Science, 22: 500 - 502

Morgan, T. H.1910. *Sex - limited inheritance in Drosophila*. Science, 32:120 - 122

———.1911. *An attempt to analyze the constitution of the chromosomes on the basis of sex - limited inheritance in Drosophila*. J. Exp. Zool. 11: 365 - 414

Bridges, B. B. (1916). Nondisjunction as a proof of the chromosome theory of heredity. Genetics 1: 1 ~ 52, 107 ~ 163

———. (1925). *Sex in relation to chromosomes and genes*. Am. Natrualist 59: 127 ~ 137

Levitan, M., and Montagu, A. (1971). *Textbook of human genetics*. New York: Oxford University Press

Hodgkin, J. (1980). *More sex ~ determination mutants of Caenorhabditis elegans*. Genetics 96: 649 ~ 664

Eicher, E. M., and Washburn, L. L. (1986). *Genetic control of primary sex determination in mice*. *Annu*. Rev. Genet. 20: 327 ~ 360

Baker, B., Nagoshi, R. N., and Burtin, K. C. (1987). *Molecular genetic aspects of sex determination in Drosophila*. BioEssays 6:66 ~ 70

Boggs, R. T., Gregor, P., Idriss, S., Belote, J. M., and Mckeown, M. (1987). *Regulation of sexual differentiation in Drosophila melanogaster via alternative splicing of RNA from the transformer*. Cell 50: 739 ~ 747

Hodgkin, J. (1989). *Drosophila sex determination: a cascade of regulated splicing*. Cell 56: 905 ~ 906

Palmer, M. S., Sinclair, A. H., Berta, P., Ellis, N. A., Goodfellow, P. N., Abbas, N. E., and fellous, M. (1990). *Genetic evidence that ZFY is not the testis ~ determining factor*. Nautre 342: 937 ~ 939

第四章　孟德尔分析的扩展

Searle, A. G. (1968). *Comparative Genetics of Coat Color in Mammals*. New York: Academic Press. A classic treatment of the subject with many examples relevant to this and other chapters

Bodmer, W. F., and L. L. Cavalli ~ Sforza. (1976). *Genetics, Evolution, and Man*. New York: W. H. Freeman and Co. A very readable, well ~ illustrated book, including a clear account of HLA genetics

Hutt, W. B. (1979). *Genetics for Dog Breeders*. New York, W. H. Freeman and Co. A short book that will make genetics immediately relevant to all dog owners and breeders

Silvers, W. K. (1979). *The Coat Colors of Mice*. New York: Springer ~ Verlag. A standard handbook in the subject, including many examples of gene interaction.

Wright, M., and S. Walters, eds. (1981). *The book of the Cat*. New York: Summit Books. A fascinating book that has an excellent chapter on gene interaction in determining the coat colors of domestic cats

Griffiths, A. J. E., and F. R. Ganders. (1984). *Wildfower Genetics*. Vancouver: Flight Press. A field guide to plant variation in natural populations and its genetic basis, including examples relevant to this chapter

Griffiths, A. J. F., and J. Mcpherson. (1989). *One hundred + Principles of Genetics*. New York: W. H. Freeman and Co. A short book that distills genetics down to its basic key concepts; useful for rapid review of the whole subject

第五章　基因的连锁与交换和真核生物的基因作图

Morgan, T. H. (1911). 孟德尔遗传的随机分离与相引. 1987. 载:G. 孟德尔等著(梁宏,王斌译)遗传学经典论文集. P55. 北京:科学出版社

Creighton, H. B. and McClintick (1931). 玉米细胞学和遗传学交换的关系. 载:G. 孟德尔等著(梁宏,王斌译). 1987. 遗传学经典论文集. P84. 北京:科学出版社

Morgan, T. H. (1910). *Sex ~ limited inheritance in Drosophila* Science 32: 120 ~ 122

———. (1910). *The method of inheritance of two sex ~ limited characters in the same animal*. Proc. Soc. Exp. Biol. Med. 8: 17

———. (1911). *An attempt to analyze the constitution of the chromosomes on the basis of sex ~ limited inheritance in Drosophila*. J. Exp. Zool. 11: 365 ~ 414

Blixt, S. (1975). *Why didn't Mendel find linkage?* Nature 256: 206

第六章　数量遗传

Nilsson ~ Ehle, H. (1909). *Kreuzungsuntersuchungen and Hafer und Weizen. Lunds Univ*. Aarskr. N. F. Atd., Ser. 2. 5(2): 1 ~ 122

East, E. M. (1910). *A Mendelian interpretation of variation that is apparently continuous*. Am. Naturalist 44: 65 ~ 82

Mather, K. (1943). *Polygenic inheritance and natural selection. Biol. Rev.* 18: 32 ~ 64 Blumer, M. G. (1980). *The mathematical theory of quantitative genetics*. Oxford: Clarendon Press

Sokal, R. R., and Rohlf, F. J. (1981). *Biometry*, 2nd ed. San Francisco: W. H. Freeman

Hill, W. G., ed (1984). *Quantitative genetics*, Parts I and II. New Yorks: Van Nostrand Reinhold Company, Inc

Paterson, A. H., Lander, E. S., Hewitt, J. D., Person, S., Lincoln, S. E., and Tanksley, S. D.(1988). *Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map*. Nature 335: 721~726

Weir, B. S., Eisen, E. J., Goodman, M. M., and Namkoong, G., eds. (1988). *Proceedings of the second international conference on quantitative genetics*. Sunderland, MA: Sinauer

第七章 基因结构和功能

Benzer,S(1955). 细菌噬菌体遗传区的精细结构. 载:G. 孟德尔等著(梁宏,王斌译).1987. 遗传学经典论文集, P55. 北京:科学出版社

Stadler,L.J.(1954). 基因. 载:G. 孟德尔等著(梁宏,王斌译).1987. 遗传学经典论文集,P55. 北京:科学出版社

Beadle, G. W., and Ephrussi, B. (1937). *Development of eye colors in Drosophila: diffusible substances and their interrelationships*. Genetics 22:76~86

———, and Tatum, E. L. (1942). *Genetic control of biochemical reactions in Neurospora*. Proc. Natl. Acad. Sci. USA 27: 499~506

Harris, H. (1975). *The principles of human biochemical genetics*. Amsterdam: North~Holland

Maniatis, T., Fritsch, E. F., Lauer, J., and Lawn, R. M.(1980). *The molecular genetics of human hemoglobins*. Annu. Rev. Genet. 14: 145~178

Scriver, C. R., and Clow, C. L. (1980). *Phenylketonuria and other phenylalanine hydroxylation mutants in man*. *Annu*. Rev. Genet. 14: 179~202

Woo, S. L. C., Lidsky, A. S., Guttler, F., Chandra, T., and Robson, K. J. H. (1983). *Cloned human phenylalanine hydroxylase gene allows prenatal diagonosis and carrier detection of classical phenylketonuria*. Nature 300: 151~155

Guljaard, H. (1986). *Biochemical diagnosis of genetic diseases*. Experientia 42L 1075~1085

Haseltine,W.A.(1997). *Discovering genes for new medicines*. Sci. Am.276(3),78~83

第八章 细菌和噬菌体的遗传重组

Jacob, F., and Wollman, E. L.(1951). *Sexuality and the genetics of bacteria*. New York: Academic Press

Benzer, S. (1959). *On the topology of the genetic fine structure*. Proc. Natl. Acad. Sci. USA 45: 1607~1620

———,(1962). *The fine stricture of the gene*. Sci. Am. 206: 70~84

Wollman, E. L., Jacob, F., and Hayes, W. (1962). *Conjugation and genetic recombination in E. coli K~12*. Cold Spring Harbor Symp. Quant. Biol. 21: 141~162

Hayes, W. (1968). *The genetics of bacteria and their viruses*, 2nd ed. New York: Wiley

Campbell, A. (1969). *Episomes*. New York: Harper& Row

Crutiss, R. (1969). *Bacterial conjugation*. Annu. Rev. Microbiol. 23: 69~136

Hotchkiss, R. D., and Gabor, M. (1970) . *Bacterial transformation with special reference to recombination processes*. Annu. Rev. Genet. 4: 193~224

Archer, L. J. (1973). *Bacterial transformation*. New York: Academic Press

Ippen~Ihler,K.A. and Minkley,E.G.(1986). *The conjugation system of F, the fertility factor of E. coli*, Ann. Rev. Genet. 20,593~624

Frost,L.S.,Ippen~Ihler,K., and Skurray, R.A.(1994). *Analysis of the sequence and gene products of the transfer region of the F sex factor*. Microbiol. Rev. 58,162~210

Lanka,E. and Wilkins,B.M.(1995). *DNA processing reactions in bacterial conjugation*. Ann.Rev. Biochem. 64, 141~169

第九章 遗传物质的分子结构和性质

Watson,J.D.& Crick,F.H.C. (1953). 核酸的分子结构. 载:G. 孟德尔等著(梁宏,王斌译).1984. 遗传学经典论文集 .P144. 北京:科学出版社

Watson,J.D.& Crick,F.H.C.(1953). 脱氧核糖核酸的结构. 载:G.孟德尔等著(梁宏,王斌译).1984. 遗传学经典论文集 .P146. 北京:科学出版社

J.D. 沃森著 (刘望夷等译).1984. 双螺旋——发现 DNA 结构的故事. 北京:科学出版社

Chargaff, E. (1951). *Structure and function of nucleic acids as cell constituents*. Fed. Proc. 10: 654 ~ 659

Watson, J. D. (1968). *The Double Helix*. New York: Atheneum. An enjoyable personal account of Watson and Crick's discovery, including the human dramas involved

Davidson, J. N. (1972). *The biochemistry of the nucleic acids*, 8th ed. New York: Academic Press

Wang, J. C. (1982). *DNA Toposiomerases*, Scientific American(July). Diagrams different topological forms of DNA

Dickerson, R. E. (1983). *The DNA Helix and How It is Read*, Scientific American(December). An article with some beautiful color models of DNA structures

Geis, I. (1983). *Visualizing the anatomy of A, B and Z - DNAs*. J. Biomol. Struct. Dynam. 1: 581 - 591

Wu, H. - M., and Crothers, D. M. (1984). *The locus of sequence - directed and protein - induced DNA bending*. Nature 308: 509 ~ 513

Rich, A., Nordheim, A., and Wang, A. H. - J. (1984). *The chemistry and biology of left - handed Z - DNA*. Annu. Rev. Biochem. 53: 791 - 846

Azorin, F., and Rich, A. (1985). *Isolation of Z - DNA binding proteins from SV40 minichromosomes:* Evidence for binding to the vital control region. Cell 41: 365 ~ 374

Wittig, B., Dorbic, T., and Rich, A. (1989). *The level of Z ~ DNA in metabolically active, permeabilized mammalian cell nuclei is regulated by torsional strain*. J. Cell. Biol. 108: 755 ~ 764

Kornberg, A., and T. Baker. (1992). *DNA Replication*, 2d ed. New York: W. H. Freeman and Co

第十章　基因组和染色体

Britten, R. J., and Kohne, D. E. (1968). *Repeated sequences in DNA*. Science 161: 529 ~ 540

Long, E. O., and David, I. B. (1980). *Repeated genes in eukaryotes*. *Annu. Rev. Biochem.* 49: 727 ~ 764

Singer, M.F. (1982). *SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes*. Cell 28: 133 ~ 134

Blackburn, E. H., and Szostak, J. W. (1984). *The molecular struture of centromeres and telomeres*. *Annu.* Rev. Biochem. 53: 163 ~ 194

Richard, T. J., finch, J. T., Rushton, B., Rhodes, D., and Klug, A. (1984). *Structure of the nucleosoem core particle at 7A resolution*. Nature 311: 532 ~ 537

Woodcock, C. L. F., Frado, L. - L. Y., and Rattner, J. B. (1984). *The higher ~ order structure of chromatin: Evidence for a helical ribbon arrangement*. J. Cell. Biol. 99: 42 ~ 52

D'Ambrosio, E., Waitzikin, S. D., Whitney, F. R., Salemme, A., and Furano, A. V. (1985). *Structure of the highly repeated, long interspersed DNA family(LINE or L1Rn) of the rat*. Mol. Cell. Biol. 6: 411 ~ 424

Singer, M.F and Skowronski, J. (1985). *Making sense out of LINES: Long interspersed repeat sequences in mammalian genomes*. Trends Biochem. Sci. (Mar.): 119 ~ 121

Burlingame, R. W., Love, W. E., Wang, B. ~ C., Hamlin, R., Xuang, N. ~ H., and Moudranakis, E. N.(1985). *Crystallographic structure of the octameric histone core of the nucleosome at a resolution of* 3.3 *A*. Science 228: 546 ~ 553

Korenberg, J. R., and Rykowski, m. C. (1988). *Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands*. Cell 53: 391 ~ 400

Amati, B. B., and Gasser, S. M.(1988). *Chromosomal ARS and CEN elements bind specifically to the yeast nuclear scaffold*. Cell 54: 967 ~ 978

Boy de la Tour, E., and Laemmli, U. K.(1988). *The metaphase scaffold is helically folded: sister chromatids have predominantly oppsite helical handedness*. Cell 55: 937 ~ 944

Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989). *Monovalent cation ~ induced struture of telomeric DNA: the G ~ quartet model*. Cell 59: 871 ~ 880

Brock, T. D. (1988). *The bacterial nucleus: a history*. Microbiol. Rev. 52, 397 ~ 411

Cai, M., and Davis, R. W. (1990). *Yeast centromere binding protein CBF1, of the helix ~ loop ~ helix protein family, is required for chromosome stability and methionine prototrophy*. Cell 61: 437 ~ 446

Schulman. I. And Bloom, K. S. (1991). *Centromeres: an integrated protein/ DNA complex required for chromosome movement*. Ann. Rev. Cell Biol. 7, 311 ~ 336

Blackbrun, E. H. (1991). *Structure and function of telomeres*. Nature 350, 569 ~ 573

Hyman, A. A. and Sorger, P. K. (1995). *Structure and function of kinetochores in budding yeast*. Ann. Rev. Cell Biol. 11, 471 ~ 495

Zakian, V. A. (1995). *Telomeres: beginning to understand the end*. Science 270, 1601 ~ 1607

Greider, C. W. (1996). *Telomere length regulation*. Ann. Rev. Biochem. 65, 337 ~ 366

Kippling, D. and Warburton, P.E. (1997). Centromeres , *CEN ~ P and Tigger too*. Trends Genet. 13: 141 ~ 144

第十一章 DNA 的复制

Scott, J. F. Et al. (1977). *A mechanism of duplex DNA replication revealed by enzymatic studies of phage ΦX174: catalytic strand separation in advance of replication*. Proc. Nat. Acad. Sci. USA 74, 193 ~ 197

Nossal, N. G. (1983). *Prokaryotic DNA replication systems*. Ann. Rev. Biochem. 52, 581 ~ 615

Campbell, J. L. (1986). *Eukaryotic DNA replication*. Ann. Rev. Biochem. 55, 733 ~ 771

Baker, T. A. and Kornberg, A. (1988). *Transcriptional activation of initiation of replication from the E. coli chromosomal origin: an RNA ~ DNA hybrid near oriC*. Cell 55, 113 ~ 123

Ogden, G. B., Pratt, M. J., and Schaechter, M. (1988). *Duplex opening by dna protein at novel sequences in initiation of replication at the origin of the E. coli chromosome*. Cell 52, 743 ~ 755

Ogden, G. B., Pratt, M. J., and Schaechter, M. (1988). *The replicative origin of the E. coli chromosome binds to cell membranes only when hemimethylated*. Cell 54, 127 ~ 135

Nchenry, C. S. (1988). *DNA polymerase III holoenzyme of E. coli*. Ann. Rev. Biochem. 57, 519 ~ 550

Stillman, B. (1989). *Initiation of eukaryotic DNA replication in vitro*. Ann. Rev. Cell. Biol. 5, 197 ~ 245

Hill, T. M. (1992). *Arrest of bacterial DNA replication*. Ann. Rev. Microbiol. 46, 603 ~ 633

Kornberg, A. and Baker, T.A. (1992). *DNA replication*. (Freeman and Co., New York)

Marians, K. J. (1992). *Prokaryotic DNA replication*. Ann. Rev. Biochem. 61, 673 ~ 719

Waga, S. And Stillman, B. (1994). *Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication in vitro*. Nature 369, 207 ~ 212

Naktinis, V., Turner, J., and O'Donnell, M. (1996). *A molecular switch in a replication machine defined by and internal competition for protein rings*. Cell 84, 137 ~ 145

Grieder, C.W., Blackburn, G.H. (1996), *Telomeres, telomerase and cancer*. Sci. Am274 (2), 80 ~ 85

Foiani, M., Lucchini, G. and Plevani, P. (1997), *The DNA polymerase α ~ primase complex couples DNAreplication, cell cycle progression and DNA damage response*. Trends Biochem. Sci .22 (11), 424 ~ 427

第十二章 转录

Reznikoff, W. S. et al. (1985). *The regulation of transcription initiation in bacteria*. Ann. Rev. Genet. 19, 355 ~ 387

Doi, R. H. and Wang, L. ~ F. (1986). *Multiple prokaryotic RNA polymerase sigma factors*. Microbiol. Rev. 50, 227 ~ 243

Platt, T. (1986). *Transcription termination and the regulation of gene expression*. Ann. Rev. Biochem. 55, 339 ~ 372

Friedman, D. I., Imperale, M. J., and Adhya, S. L. (1987). *RNA 3' end formation in the control of gene expression*. Ann. Rev. Genet. 21, 453 ~ 488

Das, A. (1993). *Control of transcription termination by RNA ~ binding proteins*. Ann. Rev.Biochem. 62, 893 ~ 930

Haldenwang, W. G. (1995). *The sigma factors of B. Subtilis*. Microbiol. Rev. 59, 1 ~ 30

Travers, A. A. and Burgess, R. R. (1969). *Cyclic reuse of the RNA polymerase sigma factor*. Nature 222, 537 ~ 540

Grossman, A. D., Erickson, J. W., and Gross, C.A. (1984). *The htpR gene product of E. coli is a sigma factor for heat ~ shock promoters*. Cell 38, 383 ~ 390

Siebenlist, U., Simpson, R. B., and Bilbert, W. (1980). *E.coli RNA polymerase interacts homologously with two different promters*. Cell 20, 269 ~ 281

Rice, G. A., Kane, C. M., and Chamberlin, M. (1991). *Footprinting analysis of mammalian RNA polymerase II along its transcript: an alternative view of transcription elongation*. Proc. Nat. Acad.Sci. USA 88, 4245 ~ 4249

Roberts, J.W. (1969). *Termination factor for RNA synthesis*. Nature 224, 1168 ~ 1174

Trends in Biochemical Sciences. Vol.21, No.9, (1996) . *Whole issue devoted to articles on RNA Polymerase Ⅱ and*

Control of Transcription

White, R. J. (1997). *Regulation of RNA polymerases Ⅰ and Ⅲ by the retinoblastoma protin: a mechanism for growth control*. Trends Biochem. Sci. 22 (3), 77 ~ 80

Reeder, R. T. and Lang, W. H. (1997), *Terminating transcription in eukaryotes: lessons learned from RNA polymerase Ⅰ*. Trends Biochem. Sci. 22 (12), 473 ~ 477

第十三章　转录后加工

Carignani, G. et al. (1983). *An RNA maturase is encoded by the first intron of the miltochondrial gene for the subunit I of cytochrome oxidase in S. cerevisiae*. Cell 35, 733 ~ 742

Cech, T. R. (1985). *Self ~ splicing RNA: implications for evolution*. Int. Rev. Cytol. 93, 3 ~ 22

Forster, A. C. and Symons, R. H. (1987). *Self ~ cleavage of virusoid RNA is performed by the proposed 55 ~ nucleotide active site*. Cell 50, 9 ~ 16

Feagin, J. E, Abraham, J. M, and Stuart, K. (1988). *Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei*. Cell 53, 413 ~ 422

Cech, T. R. (1990). *Self ~ splicing of group I introns*. Ann. Rev. Biochem. 59, 543 ~ 568

Symons, R. H. (1992). *Small catalytic RNAs*. Ann. Rev. Biochem. 61, 641 ~ 71

Lamlowitz, A. M. And belfort, M. (1993). *Introns as mobile genetic elements*. Ann. Rev. Biochem. 62, 587 ~ 622

Michel, F. And Ferat, J. ~ L. (1995). *Structure and activities of group II introns*. Ann. Rev. Biochem. 64, 435 ~ 461

Zimmerly, S. et al. (1995). *Group II intron mobility occurs by target DNA ~ primed reverse transcription*. Cell 82, 545 ~ 554

Scott, W. G, Finch, J. T, and Klug, A. (1995). *The crystal structure of an all ~ RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage*. Cell 81, 991 ~ 1002

Navaratnam, N. et al. (1995). *Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA ~ binding motif at its active site*. Cell 81, 187 ~ 195

Scott, W. G. and Klug, A. (1996), *Ribozymes: structure and mechanism in RNA catalysis*. Trends Biochem. Sci. 21 (6), 220 ~ 224

Tam, W. Y. and Steitz, J. A. (1997), *Pre ~ mRNA splicing: the discovery of a new spliceosome doubles the challenge*. Trends Biochem. Sci. 22 (4), 132 ~ 137

Kable, M. L, Heidmann, S. and Stuart, K. D. (1997). *RNA editing: getting U into RNA*. Trends Biochem. Sci. 22 (5), 162 ~ 166

第十四章　遗传密码和遗传信息的翻译系统

Holley, R. W. et al. (1965). *Structure of an RNA*. Science 147, 1462 ~ 1465

Altman, S. (1978). *Transfer RNA*. (MIT Press, Cambridge)

Rould, M. A. et al. (1989). *Structure of E. Coli glutaminyl ~ tRNA synthetase complexed with tRNAGln and ATP at 2.8Å resolution*. Science 246, 1135 ~ 1142

Ruff, M. et al. (1991). *Class II aminoacyl tRNA synthetases: crystal structure of yeast aspartyl ~ tRNA synthetase complexes with tRNA^{Asp}*. Science 252, 1682 ~ 1689

Bock, A. et al. (1991). *Selenoprotein synthesis: an expansion of the genetic code*. Trends Biochem. Sci. 16, 463 ~ 467

Jakubowski, H. and Goldman, E. (1992). *Editing of errors in selection of amino acids for protein synthesis*. Microbiol. Rev. 56, 412 ~ 429

Soll, D. And RajBhandary, U. L. (1995). *tRNA. Structure, Biosynthesis, and Function*, American Society for Microbiology, Washington DC

Farabaugh, P. J. (1996). *Programmed translational frameshifting*. Microbiol. Rev. 60, 103 ~ 134

Cedergren, R. and Miramontes, P. (1996), *The puzzling origin of the genetic code*. Trends Biochem. Sci. 21 (6), 199 ~ 200

Ramakrishnan, V. And White, S. W. (1998) *ribosomal protein structures: insights into the architecture, machinery and* evolution of the ribosome. Trends Biochem. Sci. 23 (6), 208 ~ 212

第十五章 蛋白质的合成,转运和加工

Blobel, G. And Dobberstein, B. (1975). *Transfer of proteins across the membrane. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrance-bound ribosomes of murine myeloma.* J. Cell Biol. 67, 835-851

Hautl, F.-U. et al. (1988). *Successive translocation into and out of the mitochondrial matrix: targeting of protein to intermembrane space by bipartite signal peptide.* Cell 51, 1027-1037

Collier, D. N. et al. (1988). *The antifolding activity of SecB promotes the export of the E.coli maltose-binding protein.* Cell 53, 273-283

Crooke, E. et al. (1988). *ProOmpA is stabilized for membrane translocation by either purified E. coli trigger factor or canine signal recognition particle.* Cell 54, 1003-1011

Brundage, L. et al. (1990). *The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation.* Cell 62, 649-657

Gorlich, D. and Rapoport, T. A. (1993). *Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane.* Cell 75, 615-630

Moroianu, J., Blobel, G., and Radu, A. (1995). *Previously identified protein of uncertain fuinction is importin α and together with importin β docks import substrate at nuclear pore complexes.* Proc. Nat. Acad. Sci. USA 92, 2008-2011

Gorlich, D. And Mattaj, I. W. (1996). *Nucleocytoplasmic transport.* Science 271, 1513-1518

Schatz, G. And Dobberstein, B. (1996). *Common principles of protein translocation across membranes.* Science 271, 1519-1526

Rapoport, T. A., Jungnickel, B., and Kutay, U. (1996). *Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes.* Ann. Rev. Biochem. 65, 271-303

第十六章 原核生物基因表达的调控

Jacob, F. and Monod, J. (1961). *Genetic regulatroy mechanisms in the synthesis of proteins.* J. Mol. Biol. 3, 318-356

Gilbert, W. and Muller-Hill, B. (1966). *Isolation of the lac repressor.* Proc. Nat. Acad. Sci. USA 56, 1891- 1898

Lee, F. and Yanofsky, C. (1977). *Transcription termination at the trp operon attenuators of E. coli and S. typhimurium: RNA secondary structure and regulation of termination.* Proc. Nat. Acad. Sci. USA 74, 4365-4368

Yanofsky, C. (1981). *Attenuation in the control of expression of bacterial operons.* Nature 289, 751-758

Baughman, G. And Nomura, M. (1983). *Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in E. coli.* Cell 34, 969-988

Izant, J. G. And Weintraub, H. (1984). *Inhibition of thymidine kinase gene expression by antisense RNA: amolecular approach to genetic analysis.* Cell 36, 1007-1015

Green, J., Pines, O., and Inouye, M. (1986). *The role of antisense RNA in gene regulation.* Ann. Rev. Biochem. 50, 569-597

Landick, R. And Yanofsky, C. (1987). *Transcription attenuation. In E. coli and S. typhimurium.* Ed. F. C. Neidhardt, American Society for Microbiology, Washington DC, 1276-1301

Gold, L. (1988). *Postranscriptional regulatory mechanisms in E. coli.* Ann. Rev.Biochem. 57, 199-223

Gaston, K. A. et al. (1990). *Stringent spacing requirements for transcription activation by CRP.* Cell 62, 733-743

Condon, C., Squires, C., and Squires, C. L. (1995). *Control of rRNA transcription in E. coli.* Microbiol. Rev. 59, 623-645

Friedman, A. M., Fischmann, T. O., and Steitz, T. A.,(1995). *Crystal structure of lac repressor core tetramer and its implications for DNA looping.* Science 268, 1721-1727

Lewis, M. et al. (1996). *Crystal structure of the lactose operon repressor and its complexes with DNA and inducer.* Science 271, 1247-1254

Hildebrandt, E. R. et al. (1995). *Comparison of recombination in vitro and in E. coli cells: measure of the effective concentration of DNA in vivo.* Cell 81, 331-340

第十七章　噬菌体的裂解途径和溶原化途径的选择和调控

Ptashne, M. (1967). *Isolation of the γ phage repressor*. Proc. Nat. Acad. Sci. USA 57, 306 ~ 313

Ptashne, M. (1967). *Specfic binding of the γ phage repressor to DNA*. Nature 214, 232 ~ 234

Johnson, A. D., Meyer, B. J., and Ptashne, M. (1979). *Interactions between DNA ~ bound repressors govern regulation by the phage λ repressor*. Proc. Nat. Acad. Sci. USA 75, 5061 ~ 5065

Pabo, C. O. And Lewis, M. (1982). *The operator ~ binding domain of λ repressor: structure and DNA recognition*. Nature 298, 443 ~ 447

Sauer, R. T. et al (1982). *Homology among DNA ~ binding proteins suggests use of a conserved super ~ secondary structure*. Nature 298, 447 ~ 451

Friedman, D. I. And Gottesman, M. (1982). *Lytic mode of lambda development*. *In Lambda II*. Eds. R. W. Hendrix, J. W. Roberts, F. W. Stahl, and R. A. Weisberg, Cold Spring Harbor Laboratory, New York, 21 ~ 51

Wharton, R. L., Brown, E. L., and Ptashne, M. (1984). *Substituting an α ~ helix switches the sequence specific DNA interactions of a repressor*. Cell 38, 361 ~ 369

Brennan, R. G. et al. (1990). *Protein ~ DNA conformational changes in the crystal structure of a λ Cro ~ operator complex*. Proc. Nat. Acad. Sci. USA 87, 8165 ~ 8169

Ptashne, M. (1992). *A Genetic Switch*. Cell Press and Blackwell Scientific Publications, Cambridge

第十八章　真核生物基因表达的调控

Hartwell, A. and Pelc, S. (1953). *Synthesis of DNA in normal qand irradiated cells and its relation to chromosome breakage*. Heredity Suppl. 6, 261 ~ 273

Nasmyth, K. (1982). *Molecular genetics of yeast mating type*. Ann. Rev. Genet. 16, 439 ~ 500

Toneguwa, S. (1983). *Somatic generation of antibody diversity*. Nature 302, 575 ~ 581

Steinmetz, M. and Hood, L. (1983). *Genes of the MHC complex in mouse and man*. Science 222, 727 ~ 732

Nasmyth, K. And Shore, D. (1987). *Transcriptional regulation in the yeast life cycle*. Science 237, 1162 ~ 1170

Murre, C., McCaw, P. S., and Baltimore, D. (1989). *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins*. Cell 56, 777 ~ 783

Vinson, C. R., Sigler, P. B., and McKnight, S. L. (1989). *Scissors ~ grip model for DNA recognition by a family of leucine zipper proteins*. Science 246, 911 ~ 916

Grunstein, M. (1990). *Histone function in transcription*. Ann. Rev. Cell Biol. 6, 643 ~ 678

Benezra, R. et al. (1990). *The protein Id: a negative regulator of helix ~ loop ~ helix DNA binding proteins*. Cell 61, 49 ~ 59

Boyes, J. and Bird, A. (1991). *DNA methylation inhibits transcription indirectly via a methyl ~ CpG binding protein*. Cell 64, 1123 ~ 1134

Chaillet, J. R., Vogt, T. F., Beier, D. R., and Leder, P. (1991). *Parental ~ specific methylation of an imprinted transgene is established during gametogenesis and progressively changes during embryogenesis*. Cell 66, 77 ~ 83

Lassar, A. B. et al. (1991). *Functional activity of myogenic HLH proteins requires hetero ~ oligomerization with E12/E47 ~ like proteins in vivo*. Cell 66, 305 ~ 315

Wolberger, C. et al. (1991). *Crystal structure of a MATα2 homeodomain ~ operator complex suggests a general model for homeodomain ~ DNA interactions*. Cell 67, 517 ~ 528

Gellert, M. (1992). *Molecular analysis of V(D)J recombination*. Ann. Rev. Genet. 26, 425 ~ 446

Felsenfeld, G. (1992). *Chromatin as an essential part of the transcriptional mechanism*. Nature 355, 219 ~ 224

Pabo, C. T. and Sauer, R. T. (1992). *Transcription factors: structural families and principles of DNA recognition*. Ann. Rev. Biochem. 61, 1053 ~ 1095

Tamkun, J. W., Deuring, R., Scott, M. P., Kissinger, M., Pattatucci, A. M., Kaufman, T. C., and Kennison, J. A. (1992). *Brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2*. Cell 68, 561 ~ 572

Xiong, Y., Zhang, H., and Beach, D. (1993). *Subunit rearrangement of the cyclin ~ dependent kinase is associated with cellular transformation*. Genes & Dev. 7, 1572 ~ 1583

Kamakaka, R. T., Bulger, M., and Kadonaga, J. T. (1993). *Potentiation of RNA polymerase II transcription by GAL4 ~ VP16 during but not after DNA replication and chromatin assembly*. Genes Dev. 7, 1779 ~ 1795

Harrison, D. A., Gdula, D. A., Cyne, R. S., and Corces, V. G. (1993). *A leucine zipper domain of the suppressor of hairy-wing protein mediates its repressive effect on enhancer function*. Genes Dev. 7, 1966-1978

Tsukiyama, T., Becker, P. B., and Wu, C. (1994). *ATP-dependent nucleosome disruption at a heat shock promoter mediated by binding of GAGA transcription factor*. Nature 367, 525-532

Gehring, W. J. et al. (1994). *Homeodomain-DNA recognition*. Cell 78, 211-223

Fisher, R.P. and Morgan, D.O. (1994). *A novel cyclin associates with MO15/cdk7 to form the cdk-activating kinase*. Cell 78, 713-724

Nurse, P. (1994). *Ordering S phase and M phase in the cell cycle*. Cell 79, 547-550

Hayles, J. et al. (1994). *Temporal order of S phase and mitosis in fission yeast is determined by the state of the $p34^{cdc2}$ mitotic B cyclin complex*. Cell 78, 813-822

Tsai, M.-J. and O'Malley, B. W. (1994). *Molecular mechanisms of action of steroid/thyroid receptor superfamily members*. Ann. Rev. Biochem. 63, 451-486

Rastinejad, F., Perlmann, T., Evans, R. M., and Sigler, P. B. (1995). *Structural determinants of nuclear receptor assembly on DNA direct repeats*. Nature 375, 203-211

Sheer, C.J. and Roberts, J.M. (1995). *Inhibitors of mammalian G1 cyclin-dependent kinases*. Genes & Dev. 9, 1149-1163

Weinberg, R.A. (1995). *The retinoblastoma protein and cell cycle control*. Cell 81, 323-330

Jeffrey, P.D. et al. (1995). *Mechanism of cdk activation revealed by the structure of a cyclin A-cdk2 complex*. Nature 376, 313-320

Kamakaka, R.T. (1997), *Silencers and locus control regins: opposite sides of the same coin*. Trends Biochem. Sci. 22 (4), 124-127

第十九章 癌基因与癌

Brugge, J. and Erikson, R.L. (1977). *Identification of a transformation-specific antigen induced by an avian sarcoma virus*. Nature 269, 346-348

Linzer, D. I. H. and Levine, A. J. (1979). *Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells*. Cell 14, 43-52

Waterfield, M.D. et al (1983). *Platelet derived growth factors is structurally related to the putative transforming protein $p28^{sis}$ of simain sarcoma virus*. Nature 304, 35-39

Bishop, J.M. (1983). *Cellular oncogenes and retroviruses*. Ann. Rev. Biochem. 52, 301-354

Stewart, T.A., Pattengale, P.K., and Leder, P. (1984). *Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/myc fusion genes*. Cell 38, 627-637

Varmus, H. (1984). *The molecular genetics of cellular oncogenes*. Ann. Rev. Genet. 18, 553-612

Barbacid, M (1987). *Ras genes*. Ann. Rev. Biochem. 56, 779-827

Haluska, F.G., Tsujimoto, Y., and Croce, C.M. (1987). *Oncogene activation by chromosome translocation in human malignancy*. Ann. Rev. Genet. 21, 321-335

Vaux, D.L., Cory, S., and Adams, J.M. (1988). *bcl-2 gene promotes hematopoietic cell survival and cooperates with c-myc to immortalize pre-B cells*. Nature 335, 440-442

Marshall, C.J. (1991). *Tumor suppressor genes*. Cell 64, 313-326

Howe, L. R., Leevers, S. J., Gomez, N., Nakielny, S., Cohen, P., and Marshall, C. J. (1992). *Activation of the MAP kinase pathway by the protein kinase raf*. Cell 71, 335-342

Seto, E., Usheva, A., Zambetti, G. P., Momand, J., Horikoshi, N., Weinmann, R., Levine, A. J., and shenk, T. (1992). *Wild-type p53 binds to the TATA-binding protein and represses transcription*. Proc. Nat. Acad. Sci. USA 89, 12028-12032

Donehower, L. A. et al. (1992). *Mice deficient for p53 are developmentally normal but susceptible for spontaneous* tumors. Nature 356, 215-221

Lefell, D.J. and Brash, D.E. (1996). *Sunlight and skin cancer*. Sci. Am. 275 (1), 38-49

Jansen-Durr, P. (1998). *How viral oncogenes make the cell cycle*. Trends Genet. 14: 8-10

第二十章　核外遗传信息的传递

Lightowlers, R. N., Chinnery, D. F., Turnbull, D. M. and Howell, N. (1977). *Mammalian mitochondrial genetics: heredity, heteroplasmy and disease*. Trends Genet. 13: 450 ~ 454

Ashwell, M., and Work, T. S. 1970. *The biogenesis of mitochondria*. Annu. Rev. Biochem. 39: 251 ~ 290

Kirk, J. T. O. 1971. *Chloroplast structure and biogenesis*. Annu. Rev. Biochem. 40: 161 ~ 196

Birky, C. W. 1978. *Transmission genetics of mitochondria and chloroplasts*. Annu. Rev. Genet. 12: 471 ~ 512

Linnane, A. W., and Nagley, P. 1978. *Mitochondrial genetics in perspective: The derivation of agenetic and physical map of the yest mitochondrial genome*. Plasmid 1: 324 ~ 345

Schmidt, R. J., Richardson, C. B., Gillham, N. W., and Boynton, J. E. 1983. *Sites of synthesis of chloroplast ribosomal proteins in Chlamydomonas*. J. Cell Biol. 96: 1451 ~ 1463

Clayton, D. A. (1984). *Transcription of the mammalian mitochondrial genome*. Ann. Rev. Biochem. 53, 573 ~ 594

Attardi, G. (1985). *Animal mitochondrial DNA: an extreme example of ecnomy*. Int. Rev. Cytol. 93, 93 ~ 146

Palmer, J. D. (1985). *Comparative organization of chloroplast genomes*. Ann. Rev. Genet. 19, 423 ~ 354

Rush, M. G., and Misra, R. 1985. *Extrachromosomal DNA in eukaryotes*. Plasmid 14: 177 ~ 191

Shinozaki, K. et al. (1986). *The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression*. ENBO J. 5, 2043 ~ 2049

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987). *Mitochondrial DNA and human evolution*. Nature 324, 31 ~ 36

第二十一章　基因突变

Drake, J. W. 1970. *The Molecular Basis of Mutation*. San Francisco: Holden - Day

Ames, B. N., Durston, W. E., Yamasuki, E., and Lee, F. D. (1973). *Carcinogens are mutagens: A simple test system combining liver homogenates for activation and bacteria for detection*. Proc. Natl. Acad. Sci. USA 70: 2281 ~ 2285

Sancer, A. and Rupp, W. D. (1983). *A novel repair enzyme: uvrABC excision nuclease of E. coli cuts a DNA strand on both sides of the damaged region*. Cell 33, 249 ~ 260

Radman, M., and Wagner, R. 1986. *Mismatch repair in escherichia coli*. Anu. Rev. Genet. 20: 523 ~ 538

Modrich, P. 1987. *DNA mismatch correction*. Annu. Rev. Biochem. 56: 435 ~ 466

Cairns, J., Overbaugh, J., and Miller, S. (1988). *The origin of mutants*. Science 335: 142 ~ 145

Sancar, A. (1996). *DNA excision repair*. Ann. Rev. Biochem. 65, 43 ~ 82

Wood, R. D. (1996). *Mismatch repair in eukaryotes*. Ann. Rev. Biochem. 65, 135 ~ 168

第二十二章　染色体畸变

Muller, H. J. (1927). G. 基因的人工蜕变. - G. 孟德尔等著(梁宏, 王斌译)遗传学经典论文集. 科学出版社, 1984

Barr, M. L. and Bertram, E. G. (1949). *A morphological distinction between neurones of the male and female, and the behavior of the nucleolar satellite during accelerated nuceoprotein synthesis*. Nature 163: 676 ~ 677

Lyon, M. F. (1961). *Gene action in the X - chromosomes of the mouse (Mus. Musculus L)*. Nature 190: 372 ~ 373

Penrose, L. S., and Smith, G. F. (1966). *Down's anomaly*. Boston: Little, Brown

Bloom, A. D. (1972). *Induced chromosome aberrations in man*. Adv. Hum. Genet. 3: 99 ~ 153

Rowley, J. D. 1973. *A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining*. Nautre 243: 290 ~ 293

Auerbasch, C. (1976). *Mutation research*. *London*: Chapman and Hall

DeKlein, A., van Kessel, A. G., Grosveld, G., Bartram, C. R., Hagemeijer, A., Bootsma, D., Spurr, N. K., Heisterkamp, N., Groffen, J., and Stephenson, J. R. (1982). *A cellular oncogene is translocated to the Philadelphia chromosome in chronic myclocytic leukemia*. Nature 300: 765 ~ 767

Borst, P., and Greaves, D. R. 1987. *Programmed gene rearrangements altering gene expression*. Science 235: 658 ~ 667

第二十三章 重组和转座

Holliday, R. (1964). *A mechanism for gene conversion in fungi*. Genet. Res. 5, 282 ~ 304

Stahl, F. W. (1979). *Special sites in generalized recombiantion*. Ann. Rev. Genet. 13, 7 ~ 24

Szostak, J. W. et al. (1983). *The double ~ strand break model for DNA recombination*. Cell 33, 25 ~ 35

Boeke, J. D. et al. (1985). *Ty elements transpose through an RNA intermediate*. Cell 40, 491 ~ 500

Mount, S. M. and Rubin, G. M. (1985). *Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins*. Mol. Cell. Biol. 5, 1630 ~ 1638

Loeb, D. D. et al. (1986). *The sequence of a large L1Md element reveals a tandemly repeated 5' end and several features found in retrotransposons*. Mol. Cell. Biol. 6, 168 ~ 182

Weiner, A. M., Deininger, P. L., and efstratiadis, A. (1986). *Noviral retroposons: genes, pseudogens, and transposable elements generated by the reverse flow of genetic information*. Ann. Rev. Biochem. 55, 631 ~ 661

Cox, M. M. and Lehman, I. R. (1987). *Enzymes of general recombination*. Ann. Rev. Biochem 56, 229 ~ 262

Thompson, J. F. and Landy, A. (1989). *Regulation of phage lambda site specific recombination*. In Mobile DNA. Eds. D. E. Berg. And M. M. Howe, American Society for Microbiology, Washington DC. 1 ~ 22

Hutchison, C. A. et al. (1989). *LINEs and related retroposons: long interspersed repeated sequences in the eukaryotic genome*. In Mobile DNA. (Eds. M. M. Howe and D. E. Berg. American society for Microbiology, Washingto DC, 593 ~ 617

Deininger, P. L. (1989). SINEs: *short interspersed repeated DNA elements in higher eukaryotes*. In Nobile DNA. Eds. M. M. Howe and D. E. Berg. American Society for Microbiology, Washington DC, 619 ~ 636

Sun, H., Treco, D., and Szostak, J. W. (1991). *Extensive 3' overhanging, single ~ stranded DNA associated with the meiosis ~ specific double ~ stranded DNA associated with the meiosis ~ specific double ~ strand breaks at the ARG4 recombination initiation site*. Cell 64, 1155 ~ 1161

West, S. C. (1992). *Enzymes and molecular mechanisms of genetic recombination*. Ann. Rev. Biochem. 61, 603 ~ 640

Luan, D. D. et al. (1993). *Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non ~ LTR retrotransposition*. Cell 72, 595 ~ 605

Lauermann, V. And Boeke, J. D. (1994). *The primer tRNA sequence is not inherited during Ty1 retrotransposition*. Proc. Nat. Acad. Sci. USA 91, 9847 ~ 9851

Kowalczykowski, S. C. et al. (1994). *Biochemistry of homologous recombination in E. coli*. Microbiol. Rev. 58m 401 ~ 465

Wang, J.C. (1996). *DNA topoisomerases*. Ann. Rev. Biochem. 65, 635 ~ 692

Kodadek, T. (1998). *Mechanistic parallels between DNA replication, recombination and transcription*. Trends Biochem. 23 (2), 79 ~ 93

第二十四章 基因调控和细胞的分化发育

Nusslein ~ Vollhard, C. and Wieschaus, E. (1980). *Mutations affecting segment number and polarity in Drosophila*. Nature 287, 795 ~ 801

McGinnis, W. et al. (1984). *A homologous protein ~ coding sequence in Drosophila homeotic genes and its conservation in other metazoans*. Cell 37, 403 ~ 408

Anderson, K. V., Jungens, G., and Nusslein ~ Volhard, C. (1985). *Establishment of dorsal ~ ventral polarity in the Drosophila embryo: genetic studies on the role of the Toll gene product*. Cell 42, 779 ~ 789

Anderson, K. V., Bokla, L., and Nusslein ~ Volardm, C. (1985). *Establishment of dorsal ~ ventral polarity in the Drosophila embryo: the induction of polarity by the Toll gene product*. Cell 42, 791 ~ 798

Schupbach, T. (1987). *Germ line and soma cooperate during oogenesis to establish the dorsoventral pattern of egg shell and embryo in Drosophila mdlanogaster*. Cell 49, 699 ~ 707

Scott, M. P. And Carroll, S. B. (1987). *The segmentation and homeotic gene network in early Drosophila development*. Cell 51, 689 ~ 698

Scott, M. P. (1987). *Complex loci of Drosophila*. Ann. Rev. Biochem. 56, 195 ~ 227

Driever, W. and Nusslein ~ Volhard, C. (1988). *The bicoid protein determines position in the Drosophila embryo in a*

concentration dependent manner. Cell 54, 95 ~ 104

Ingham, P. W. (1988). *The molecular genetics of enbryonic pattern formation in Drosophila*. Nature 335, 25 ~ 34

Sinclair, A.H., Berta, P. , Palmer M.S. et al .(1990) . *A gene from the human sex ~ determining region encodes a protin with homology to a conserved DNA ~ binding motif*. Nature, 346: 240 ~ 245

St. Johnston, D. and Nusslein ~ Volhard, C. (1992) . *The origin of pattern and polarity in the* Drosophila embryo. Cell 68, 201 ~ 219

McGinnis, W. and Krumlauf, R. (1992). *Homebox genes and axial patterning*. Cell 68, 283 ~ 302

Sprenger, F. and Nusslein ~ Volhard, C. (1992). *Torso receptor activity is regulated by a diffusible ligand produced at the extracellular terminal regions of the Drosophila egg*. Cell 71, 987 ~ 1001

Krumlauf, R. (1994). *Hox genes in vertebrate development*. Cell 78, 191 ~ 202

Garcia ~ Fernandez, J. and Holland, P. W. H. (1994). *Archetypal organization of the amphioxus Hox gene cluster*. Nature 370, 563 ~ 566

Morisato, and Anderson, K.V. (1995). *Signaling pathways that establish the dorsal ~ ventral patternof the Drosophila embryo*. Ann. Rev. Genet. 29, 371 ~ 399

Martin, C. H. et al. (1995). *Complete sequence of the bithorax complex of Drosophila*. Proc. Nat. Acad. Sci. USA 92, 8398 ~ 8402

Jimenez , R., Sanchez, A., Burgos, M. et al. (1996), *Puzzing out the genetics of mammalian sex determination*. TIG. 12 (5): 164 ~ 166

第二十五章　群体遗传学

Hardy, G. H. (1908). *Mendelian proportions in a mixed population*. Science 28: 49 ~ 50

Dobzhansky, T. (1951). *Genetics and the origin of species*, 3rd ed. New York: columbia University Press

Buri, P. (1956). *Gene frequency in small populations of mutant Drosophila*. Evolution 10: 367 ~ 402

Ayala, F. J., ed. (1976). *Molecular evolution. Sunderland*, MA: Sinauer

Avise, J. C. (1986) . *Mitochondrial DNA and the evolutionary genetics of higher animals*. Phil. Trans. Roy. Soc. London, Ser. B 321: 325 ~ 342

Hartl, D.L., and Clark, A. G. (1988). *Principles of population genetics*, 2nd ed. Sunderland, MA: Sinauer

Maynard Smith, J. (1989). *Evolutionary genetics*. Oxford: Oxford University Press

Nei, M., and Koehn, R. K. 1983. *Evolution of genes and proteins*. Sunderland, MA: Sinauer

Stringer, C. B. 1990. *The emergence of modern humans*. Sci. Am. (December): 98 ~ 104

Hillis, D. M. and Moritz, C. (1990). *Molecular systematics*. Sunderland, MA: Sinauer

————. And Moritz, C., Porter, C. A., and Baker, R. J. 1991. *Evidence for biased gene conversion in concerted evolution of ribosomal DNA*. Science 251: 308 ~ 309

Cavalli ~ Sforza, L.L. (1998), *The DNA revolution in population genetics*. Trends Genet. 14: 60 ~ 65

名词解释

A　腺嘌呤(adenine)

abortive transduction　流产转导:转导的 DNA 片段末端掺入到受体的染色体中,在后代中丢失。

acentric chromosome　端着丝粒染色体:染色体的着丝粒在最末端。

Achondroplasia　软骨发育不全:人类的一种常染色体显性遗传病,表型为四肢粗短、鞍鼻、腰椎前凸。

acrocetric chromosome　近端着丝粒染色体:着丝粒位于染色体末端附近。

active site　活性位点:蛋白质结构中具有生物活性的结构域。

adapation　适应:在进化中一些生物的可遗传性状发生改变,使其在一定的环境能更好地生存和繁殖。

adenine　腺嘌呤:在 DNA 中和胸腺嘧啶配对的碱基。

albino　白化体:一种常染色体隐性遗传突变。动物或人的皮肤及毛发呈白色,主要因为在黑色素合成过程中,控制合成酪氨酸酶的基因发生突变所致。

allele　等位基因:一个座位上的基因所具有的几种不同形式之一。

allelic frequencies (one frequencies)　在群体中存在于所有个体中某一个座位上等位基因的频率。

allelic exclusion　等位排斥:杂合状态的免疫球蛋白基因座位中,只有一个基因因重排而得以表达,其等位基因不再重排而无活性。

allopolyploicly　异源多倍体:多倍体的生物中有一套或多套染色体来源于不同物种。

Ames test　埃姆斯测验法:Bruce Ames 于 1970 年人用鼠伤寒沙门氏菌(大鼠)肝微粒体法来检测某些物质是否有诱变作用。

amino acids　氨基酸:是构成蛋白质的基本单位,自然界中存在 20 种不同的氨基酸。

aminoacyl-tRNA　氨基酰 - tRNA:tRNA 的氨基臂上结合有相应的氨基酸,并将氨基酸运转到核糖体上合成蛋白质。

aminoacyl-tRNA synthetase　氨基酰 - tRNA 合成酶:催化一个特定的 tRNA 结合到相应的 tRNA 分子上。因有 20 种氨基酸,故有 20 种氨基酰 - tRNA 合成酶。

amniocentesis　羊膜穿刺术:产前诊断中一种采羊水的方法。

amorph　无效等位基因:一种突变的形式,突变后的基因不能指令合成有功能的蛋白质。

amphiodiploid　双二倍体:即异源四倍体。

amplification　扩增:许多 DNA 拷贝来自于一个 DNA 序列。

aneuploid　非整倍体:一种染色体数目的变异体,细胞中增加或减少一条或几条染色体。

annealing　退火:即 DNA 复性,降低温度使两条 DNA 单链重新互补结合形成双链的过程。

antibody　抗体:一种免疫蛋白分子,由免疫系统产生,可识别特异抗原并与之结合。

anticodon　反密码子:在 tRNA 的反密码子环上的三个相连的碱基可和 mRNA 上的密码子互补结合。

antigen　抗原:任何一种可刺激机体产生抗体,并能与之特异结合的大分子物质。

AP site　AP 位点:即脱嘧啶脱嘌呤的位点,由 DNA 上碱基糖苷链的断裂形成,在修复时可供脱嘧啶嘌呤内切酶识别。

ascospore　子囊孢子:某些真菌产生的有性孢子,位于子囊中,为单倍性,常呈线状排列。

attached X　并联 X 染色体:果蝇的一对 X 染色体端端相连,作为一个单位进行遗传。

attenuator　衰减子或弱化子:在色氨酸操纵子结构基因中的一个区域,此区域以形成不同二级结构的方式,利用原核生物转录与翻译的偶联对转录进行调节。此区域只存在于原核生物合成代谢的操纵子中。

autonomous replication sequence (ARS)　自主复制序列:酵母 DNA 中的一个特殊片段,为复制启始

所需，是复制系统蛋白识别和结合的位点。

autoplouploid　同源多倍体：通过单套基因组加倍而形成的多倍体。

autoradiography　放射自显影：在组织或 DNA 中的放射性标记使 X 胶片曝光的过程。

autoregulation　自我调节：基因通过自身的产物来调节转录。

autosome　常染色体：性染色体以外的任何染色体。

auxotroph　营养缺陷型：微生物的一种突变体，它不能合成生长所需的物质，培养时必须在培养基中加入此物质才能生长。

back mutation　回复突变：见 reversion

bacteriophage（phage）　一种感染细菌的病毒。

balance model　平衡模型：关于遗传变异比例的一种模型，它认为自然选择维持了群体中大量遗传变异的存在。

balanced polymorphism　平衡多态现象：稳定的遗传多态现象是由自然选择来维持的。

Barr body　巴氏小体：在正常雌性哺乳动物的体细胞核中有一个高度凝聚的染色质团，它是一个失活的 X 染色体。

base analog　碱基类似物：一种化学物质，其分子结构和 DNA 的碱基相似，在 DNA 的代谢过程中有时会取代正常碱基，结果使 DNA 的碱基发生突变。

bead theory　串珠学说：已被否定的学说，认为基因附着在染色体上，就像项链上的串珠。它既是突变单位又是重组单位。

binary fission　二分分裂：一个细胞分裂为大小相近的两个子细胞的过程。

binomial distribution　二项分布：具有两种可能结果的事件其概率函数的分布，是一种离散型分布。

biparental zygote　双亲合子：又称双亲遗传（biparental inheriance），衣藻（*chlamydomonas*）的合子含有来自双亲的 DNA。这种细胞一般很少见。

biochemical mutation　生化突变，见自发突变（autotrophic mutation）。

bivalent　二价体：在第一次减数分裂时彼此联合的一对同源染色体。

bottleneck effect　瓶颈效应：一种类型的漂变。当群体很小时产生这种效应，结果使基因座中有的基因丢失了。

branch-point sequence　分支点序列：在哺乳动物细胞中的保守序列：YNCURAY（Y：嘧啶，R：嘌呤，N：任何碱基），位于核 mRNA 内含子和Ⅱ类内含子 3′端附近，其中的 A 可通过 5′-2′连接的方式和内含子 5′端相连接，在剪接时形成套马索状结构。

broad-sense heritability　广义遗传力：表型方差中所含遗传方差的百分比。

cotplot　浓度时间乘积图：一个样本单位单链 DNA 分子复性动力学曲线。以结合为双链的量为纵坐标，以 DNA 浓度和时间的乘积为横坐标作出的 DNA 复性动力学曲线。

C value　C 值：生物单倍体基因所含的 DNA 总量。

CAAT element　CAAT 元件：真核启动子上游元件之一，常位于上游 -80bp 附近，其功能是控制转录起始频率，保守序列是 5′-GGCCAATCT-3′。

cancer　癌：恶性肿瘤，细胞分裂失控，异常分裂且在生物体内可扩散。

5′-capping 5′　加帽：在 mRNA 加工的过程中在前体 mRNA 分子的 5′端加上甲基核苷酸的“帽子”。

catabolite repression（glucose effect）　分解代谢物阻遏（糖效应）：当糖存在时能诱发细菌操纵子的失活，即使操纵子的诱导物存在也是如此。

cDNA　互补 DNA：以 mRNA 为模板，以反转录酶催化合成的 DNA 的拷贝。

cDNA clone　cDNA 分子克隆：将 cDNA 片段装在载体上转化细菌扩增出多克隆的过程，最终可建立 cDNA 文库。

cDNA library　cDNA 文库：收集各 cDNA 分子的克隆，各种 cDNA 分子是细胞中总体 mRNA 的拷贝。

cell division　细胞分裂：一个细胞分裂产生两个子细胞的过程。

cell-free，protein-synthesizing system　无细胞蛋白合成系统：是表达系统之一，此系统含有核糖体、tRNA 和各种氨基酸及体外合成多肽所必须的一系列蛋白因子。只要加入 mRNA 即可表达。

cellular oncogene (c-onc) 细胞癌基因：在动物细胞中存在的一些基因，它们具有正常的功能，但一旦突变会导致细胞的癌变。

centi-Morgan (cM) 厘摩：是作图单位，相当于 1 个图距单位。

centromere (kinetochore) 着丝粒（着丝点）：染色体上的特殊区域，在显微镜下可看到的一种结构。在细胞分裂时是纺锤丝附着处。

chain-teminatiog codon 链－终止密码子：三个无义密码子。蛋白质合成时核糖体阅读这些密码子时终止合成。

charged tRNA 负载 tRNA：已带有相应氨基酸的 tRNA。

chiasma 交义：在交换时一对染色体出现的交叉型结构。在有丝分裂的双线期可以在显微镜下看到。

chimere 嵌合体：一种异源嵌合体，可能由两个不同的受精卵偶尔融合而成。

chi-square test 卡方检测：一种检测理论值与观察值之间有无差异的统计方法。

chloroplast 叶绿体：植物中进行光合作用的一种细胞器，内有双链 DNA。

chromatid 染色单体：复制后两个染色体着丝粒相连，其中的任意一条称为染色单体。

chromatin 染色质：染色体的基质，它包含 DNA、染色体蛋白和染色体 RNA。

chromocenter 染色中心：多线染色体附着在一起的过程。

chromomere 染色粒：在减数分裂和有丝分裂前期可观察到染色体上的小珠状结构。

chromosome 染色体：基因端相连，线状排列，有时结合蛋白质和 RNA。

chromosome aberration 染色体畸变：染色体结构和数目的改变。

chromosomal interference 染色体干涉：由于染色单体的断裂和重接而导致的物理干涉，阻碍了第二次交换的发生。

chromosome theory of heredity 遗传的染色体学说：认为染色体是基因的载体。此学说是由 Sutton 和 Boveri 各自独立提出的。他们的依据是染色体传递和基因传递的平行性。

chromosome map 染色体图：见连锁图(*linkage map*)。

chromosome puff 染色体疏松：染色体的某些位置上产生膨松，此位置是转录活性位点。

shromosome rearrangement 染色体重排：涉及到染色体重新连接的突变。

chromosome set 染色体组：一组染色体，对于特异物种携带一套基本遗传信息。

chromosome walking 染色体步移：一种分析真核生物大片段 DNA 的方法，常用来从基因文库中筛选特定的重组 DNA 克隆。基因文库中必须包含染色体的全部 DNA 序列，这些序列是一系列有部分重叠的片断。从某个已鉴定的克隆基因或序列开始做一系列的克隆杂交，这个基因已知与目的基因位于同一染色体上。以已知基因做探针去钓出含有邻近序列的克隆，然后以邻近序列做探针去钓出与它邻近序列的另一克隆，依次类推，每次杂交就把已鉴定的基因沿染色体向前移了一段。

cis-conformation 顺式构象：在杂合子中涉及到基因中两个突变位点，它们的排列为 $a_1a_2/++$。

cis dominance 顺式显性：一个基因对同一条染色体上下一个基因有影响能力。

cis-tran test 顺反测验：即互补测验。通过测验可确定两个表型效应相似的突变位点是否位于同一个顺反子或基因内。

cis-tran 顺反子：原来定义为一个遗传的功能单位，顺反子内部的两个突变位点不能互补。现在它与基因的含义相同，是编码有功能的 RNA 分子。

classical model 经典模型：群体遗传学中对于遗传变异解释的一种学说，提出含有较少遗传变异的自然群体是对一个等位基因有很强的选择作用的结果。

clonal selection 克隆选择：特异的 B 淋巴细胞表面受到相应抗原的刺激时便大量增殖，并分泌相应的抗体，此过程称为克隆选择。

clone 克隆：(1)一群遗传上相同的细胞或由一个共同的祖先经无性分裂产生的群体。(2)一个个体通过某种无性过程而产生在遗传上相同的后代。(3)见 DNA clone 或 cloning。

cloning 克隆：DNA 分子通过在合适的宿主中复制产生很多拷贝(如一个重组 DNA 分子)。

cloning vector (cloning vehicle) 克隆载体：一种双链环状 DNA 分子，它在宿主细胞中能自主复制，并可装接外源 DNA 片段而形成重组质粒进行克隆。

clustered repeated sequence 成簇的重复序列：重复序列串联排列在一起。

coding sequence 编码序列：mRNA 分子的部分，在翻译时它对应于多肽的氨基酸序列。

code dictionary 密码辞典：一种 64 个密码子及其翻译含义的表格。

codominance 共显性：在杂合状态时，一个座位中的两个等位基因同时都显现出表型效应。

coden 密码子：在 mRNA 分子中一组三个相连的核苷酸，对应于多肽链中的氨基酸序列或决定多肽的合成。

coefficent of coincidence 并发系数：通过遗传作图来确定干涉是否存在，即用实际双交换的值除以同一区域预期的双交换的值 ($s = v/\alpha\beta$)，当 $0 < s < 1$ 为正干涉；$s = 0$ 为完全干涉；$s > 1$ 为负干涉；$s = 1$ 为无干涉。

cohensive ends 黏性末端：经 II 型限制酶切后留下的单链突出末端。相同的末端可通过形成氢键互补结合。

cointegrate 共和体：两个环状的因子融合形成一个单个的大环，这是某些转座因子转座过程中的一个步骤。

colony 菌落：平板上可见到的细胞无性系集中在一起。

competent 感受态的：细胞处于能被转化外源 DNA 的状态。

complementary DNA 互补 DNA：即 cDNA（见 cDNA）。

complementary test 互补测验：见顺反测验(cis - trans test)。

complete dominance 完全显性：当杂合体 Aa 和纯合体 AA 的表型相同时，称完全显性。

concerted evolution(molecular drive) 协同进化(分子趋动)：一种很不平衡的进化过程，在基因的多拷贝中产生了均一序列。

conditional mutation 条件突变：一种突变在一定条件下表现为野生表型，但在特殊条件下出现突变表型。

conjugation 接合：F^+ 和 F^- 的两种细菌彼此接触产生接合管并单向地 ($F^+ \rightarrow F^-$) 传递遗传物质的过程。

consensus sequence 保守序列：普遍存在的共同序列，在不同的基因中具有相同的功能，如启动子、操纵子和剪切位点等。

conservative replication 保守复制：已被否定的一种 DNA 复制模型，认为一半子代 DNA 分子具有新合成的两条互补链。

constant region 恒定区：抗体分子的恒定区，简称 c 区，不同物种的抗体此区域是基本相同的，无特异性。

constitutive gene 组成型基因：一些基因其产物维持细胞的正常功能，其表达不受环境的影响，这些基因称为组成型基因，在生长细胞中它总具有活性。

constitutive heterochromatin 组成型异染色质：在染色体上总是存在的异染色质区域，如着丝粒。

control element 调控因子：一种可移动的遗传因子，可能导致靶基因产生不稳定的突变，它们以两种形式存在，即调节子和受体元件。

copy-choice model 模板选择假说：DNA 重组机制的一种模型，提出在染色体分裂时产生交换，且只能发生在两个新合成的非姐妹染色单体之间，实验证据不支持这一模型。

core enzyme 核心酶：*E. coli* RNA 聚合酶的蛋白，具有酶的活性，可写为 $\alpha_2\beta\beta'\omega$。

cosmid 黏粒：一种克隆载体，由质粒 pBR322 和 λ 噬菌体的 cos 黏性末端构建而成，大小为 4.6kb，容量为 29 ~ 45kb。

contransduction 共转导：同时转导两个或更多的细菌基因，可很好地表明基因的紧密连锁程度。

cotransformation 共转化：同时转化两个细菌的标记基因。

coupling 相引：Bateson 和 Punnett(1906) 在研究两对性状遗传时出现连锁的现象时并未能揭示其机制，提出相引和相斥(repulsion)来解释，认为亲组合的比例高是两个基因之间相互吸引，重组合

少是两个基因之间相互排斥。现在这两个词相当于顺式(cis)和反式(tron)。

covariance 协方差:统计中用来计算两个变量之间的相关系数,其计算通过公式 $(x-\bar{x})(y-\bar{y})$,x 和 y 是变量,$\bar{x}$ 是 x 值的平均值 $\bar{y}$ 是 y 值的平均值。

cri-du-chat syndrome 猫叫综合征:人类的一种染色体病,由第 5 号染色体短臂缺失所引起。患儿的哭声像猫叫。

crisscross inheritance 交叉遗传:隐性 X 连锁的性状的基因常由父亲遗传给女儿,再由女儿遗传给外孙,这种遗传方式称交叉遗传。

cross 杂交:在遗传分析中有意识地将两个基因型不同的亲本进行交配称杂交。

cross - over 交换:Morgen 和 E.Cattell(1912)提出的术语,指在同源染色体之间由于断裂和重接而产生重组的过程。

cystic fibrosis 囊性纤维化:人类的潜在致死性分泌型疾病,最突出的症状是粘液过量分泌,为常染色体隐性遗传疾病。

cytochromes 细胞色素:线粒体膜中的一种蛋白,主要的功能是将 ADP 经氧化磷酸化变为 ATP。

cytogenetics 细胞遗传学:用细胞学的方法研究遗传学,主要用显微镜研究染色体。

cytoplasmic inheritance 细胞质遗传:细胞器基因的遗传方式。

dark repair 暗修复:即切除修复(excision repair)。

daughter chromosome 子染色体:后期(有丝分裂中)或后期 II (减数分裂)产生的染色单体。

degeneracy 简并:有的氨基酸有多个相应的密码子,这种现象称为简并。

deletion 缺失:染色体丢失了某一片段。

denaturation 变性:在一定条件下(高温或变性剂等)DNA 双螺旋解链的过程。

deoxyribonuclease 脱氧核糖核酸酶:一种催化 DNA 降解成核苷的酶,也叫 DNase。

deoxyribonucleic acid 脱氧核糖核酸:由脱氧核糖核苷酸构成的高分子聚合物。互补的双链呈双螺旋形,它是绝大多数生物的遗传物质。

deoxyribose 脱氧核糖:DNA 中 5 - 碳糖,2′ - 位置上不是 - OH 基,故称脱氧核糖。

determinant 决定簇:一种特殊的定位分子,它导致细胞发育采用了特定的命运。

determination 决定:细胞对特殊命运定型的过程。

development 发育:从一个细胞变成生物个体的过程。

diakinesis 终变期:减数分裂中的一个阶段,也叫浓缩期。细胞处于前期末,即将进入分裂期,可观察到"X"形、"O"形染色体。

dicentric chromosome 双着丝粒染色体:带有两个着丝粒的染色体,常为不平衡易位所造成,在细胞分裂时可形成染色体桥,断裂后产生进一步的重复和缺失。

dideoxy nucletide 双脱氧核苷酸:一种修饰的核苷酸,其脱氧核糖的 2′ - OH 和 3′ - OH 都因脱氧而变成 2′ - H 和 3′ - H。这样下一个核苷酸不能通过 5′磷酸与之形成 3′, 5′ - 磷酸二酯键,使 DNA 链的延伸终止在这个双脱氧核苷酸处。

dideoxy (sanger) sequencing Sanger 双脱氧测序法:一种测序方法,Sanger 在加减法原理的基础上,利用双脱氧核苷酸可终止 DNA 链的延伸的原理来测序。

differentiation 分化:在发育中细胞的形态和生理功能发生了改变,最终产生了不同的组织和器官。

dihybrid cross 双因子杂种杂交:某些类型的双因子杂种之间的杂交。双因子杂种是指个体在两个不同的基因座位上两对等位基因是杂合的。

dimorphism 双态现象:一个群体中只有两种不同类型的多态现象。

diploid 二倍体:一个细胞或个体的每个细胞中具有两套染色体。

diplonema 双线期:减数分裂中前期 I 的第四阶段。

directed mutagenesis 定向诱变:改变克隆基因的特定部分,再把被修饰的基因重新导入有机体中。

directional selection 定向选择:以一个恒定的方向进行选择、使等位基因的频率发生改变,不是使一个等位基因固定就是使其淘汰。

discontinuous DNA replication DNA 不连续复制:在半不连续复制发现以前,人们认为 DNA 的前导

链和后滞链的合成都是不连续的，即先合成短的 DNA 片段，然后连接成长片段。

dispersive replication　散乱复制：DNA 合成的散乱模型，认为在子代 DNA 分子中，亲代片段和新合成的片段随机交替排列。

DNA clone　DNA 克隆：将特定 DNA 片段插入载体分子中，然后转化细菌，使重组载体形成很多拷贝。

DNA fingerprint　DNA 指纹：用限制酶来切 VNTRs(可变数同向重复序列)，然后用 VNTR 特异探针和酶切片段进行 Southern 杂交，经放射自显影产生大量的杂交带，不同人的 VNTRs 的位置不同，因此杂交带的位置也不同，故称 DNA 指纹。

DNA helicase　DNA 解旋酶：在 *E.coli* 复制中，催化 DNA 双链解旋的一种酶，是 *rep* 基因的产物。

DNA ligase　DNA 连接酶：在 DNA 复制和修复时，催化 DNA 分子游离单链末端之间形成共价键的一种酶。

DNA polymerase　DNA 聚合酶：一种催化 DNA 合成的酶。

DNA polymerase Ⅰ　DNA 聚合酶Ⅰ：*E.coli* 中催化 DNA 合成的酶，原称 Kornberg 酶，它主要负责损伤的修复和冈崎片段之间隙的填补。

DNA primase　DNA 引发酶：在 G4 和 ΦX174 噬菌体 DNA 合成中，此酶催化合成短的核苷酸引物。

docking protein　停泊蛋白或船坞蛋白，停靠蛋白：粗面内质网(ER)的膜结合蛋白，在蛋白越膜前，结合在核糖体新生肽的信号识别蛋白(SRP)、可识别 ER 膜上的停泊蛋白使具有信号肽的多肽和 ER 结合。

dominance variance (V_D)　显性方差：在遗传方差中由基因的显性效应所产生的部分。

dominant　显性：无论在纯合还是杂合时可得到表达的等位基因或表型。

dominant lethal allele　显性等位致死：一个等位基因无论在纯合还是杂合时都产生致死效应。

doseage compensation　剂量补偿：在性别决定中 XY 型的动物的雌体和雄体的 X 染色体的数目虽然不同，但它们具有一种机制使 X 连锁基因的产物数量相等，如哺乳动物用雌性动物一条 X 染色体失活的方式来达到剂量补偿。

double crossover　双交换：在减数分裂时染色体的两个标记之间发生了两次交换。

double helix　双螺旋：1953 年，由 Watson 和 Crick 提出的 DNA 分子结构模型，两条反向平行的 DNA 单链彼此由氢键 A－T，G－C 互补连接，整个分子成双螺旋状。

double infection　双感染或并感染：用两种不同的噬菌体同时感染细菌。

Down syndrome　唐氏综合征：一种人类的染色体病，患者的第 21 号染色体多出一条，故称 21－三体病，患者发育迟缓，智力低下。

drift　漂变：即遗传漂变，由于群体太小使基因频率发生随机波动，有时会丢失某基因而固定另一基因。

Duchenne muscular dystrophy　杜兴氏肌营养不良：又称假性肥大型营养不良或进行性营养不良，是人类中一种 X 隐性伴性遗传疾病，患者肌无力，几岁时发病，十几岁夭折。

duplication　重复：染色体结构变异的一种，即染色体组多了一段染色体片段。

dyad　二联体：在第一次减数分裂时由着丝粒连接的一对姊妹染色单体。

ecdysone　蜕皮激素：昆虫中的一种蜕皮的激素。

ectopic integration　异位表达：一个基因在通常不会表达的组织中表达。可由增强子的导入而导致异位表达。

ectopic integration　异位整合：在转基因生物中，一个导入的基因插入到新的座位中。

effective poplation size　有效种群大小：成体对产生下一代所贡献的配子的有效数目。

effector　效应物：一种和很多被调节基因控制有关的小分子物质。

endogenote　内基因子：在部分二倍体中原有的染色体部分，它和外基因子同源。

eendonuclease　内切核酸酶：又称核酸内切酶，它能切开多核苷酸中磷酸二酯键。

endopoloidy　核内多倍性：由于核分裂而使染色体套数增加，但胞质不分裂，如唾液腺染色体。

endosperm　胚乳：在显花植物中胚的周围有一种三倍体的组织叫胚乳，它供给种子萌发和发育的营

养，它由花粉中两个雄核中的一个和胚囊中的两个极核结合而成。

enhancer 增强子：在真核生物中，启动子的远上游存在一种调节元件，它可增强转录，具有远距离效应和无方向性两特点，属顺式调节因子。

enhancer trap 增强子阱：插入在染色体上的一种基因结构，被用于鉴定基因组中组织特异性内含子，在这种结构中，对增强子调节敏感的启动子接上报告基因，这样报告基因的表达方式反映了被调节的情况。

environmental variance (V_E) 环境方差：在表型方差中由于环境因素产生的变异。

episome 附加体（游离体、游离基因）一种自主复制的质粒，可以游离和整合两种状态存在。

epistasis 上位性：基因相互作用的一种形式，即在一个座位上基因型表达为不同的表型取决于另一个座位上的基因。

ethidium bromide (EB) 溴化乙锭：DNA 电泳的染色剂，它是一种扁平的分子，可插入 DNA 双链中，在紫外线的照射下会发出橙红色荧光。EB 对人有毒害。

euchromatin 常染色质：在分裂期发生凝聚，而在间期不凝聚的染色质。

eugenics 优生学：一门利用遗传学原理通过控制婚配来提高人类素质，减少遗传疾病的学科。

eukaryote 真核生物：细胞核有核膜的生物，其遗传物质集中在核内。真核生物有单细胞的（如真菌），也有多细胞的（如高等生物）。

eupliod 整倍体：具有完整套染色体的生物，可是单套、两套或多套。

excision repair (dark repair) 切除修复（暗修复）：不依赖光的酶催化修复 DNA 的过程。紫外线的照射可导致 DNA 产生胸苷二聚体，影响其复制，修复酶可切除胸苷二聚体而予以修复。

excojugant 接合后体：细菌接合后 F^- 细胞含有供体（F^+）细胞的 DNA 片段，称接合后体。

exon 外显子：一个基因不包括内含子的任何编码序列，其 mRNA 的相应区域可翻译成蛋白。

exonnuclease 核酸外切酶：当核苷酸链存在游离端时，这种酶可降解 DNA 分子。

expressivity 表现度：特定的基因型表达成表型的程度。

F^- cell F^- 细胞：*E. coli* K 菌株中无 F 因子的细胞。

F^+ cell F^+ 细胞：*E. coli* K 菌株中具有 F 因子的细胞。

F factor F 因子：即可育因子（fertility factor），是细菌的一种附加因子，它的存在使宿主具有供体的能力。

F′factor F′ 因子：在 F 因子上带有宿主的染色体片段，因此可将特定的基因以很高的频率传递给受体，产生性导。

F_1 generation 杂种 1 代：由两个基因型不同的亲本杂交产生的第一代杂种。

F_2 generation 杂种 2 代：由 F_1 自交或互交产生的第二代杂种。

F-pili (sexpili) F 伞毛（性伞毛）：是 F^+ 细胞中的 F 因子合成的一种很细的管状蛋白，位于细胞表面，可使 F^+ 和 F^- 或 Hfr 品系和 F^- 细胞彼此连接。

facultative heterochromatin 兼性异染色质：某些染色质在某一时期或特定的条件下会发生凝聚而失活，但它本身是含有基因的，如雌性哺乳动物两个 X 染色体中的一个 X 染色体部分失活。

familial trait 家族性状：由一个家族的成员显示出的性状，有的是遗传的，有的是环境造成的，如饮食造成的肥胖。

fine-structure mapping 精细结构作图：由 Benzer.S 建立的在一个基因内部不同等位位点之间的高分辨作图。

first law 第一定律（孟德尔的分离律）在配子形成时成对的等位基因彼此分离，并独立地分配到不同的性细胞中。

fitness 适合度：在一定的环境中一种基因型的相对繁殖能力。

fixed breakage point 固定段裂点：根据异源 DNA 重组模型，DNA 双螺旋从这一位点开始解旋，作为形成异源 DNA 的开头。

fluctuation test 波动检测法：一种用于检测微生物随机突变的性质或计算突变率的方法。

formylmethonine (fMet) 甲酰甲硫氨酸：一种在甲硫氨酸上加上甲酰基的特殊修饰氨基酸，是原核

生物和真核生物细胞器中多肽合成的第一个氨基酸，合成开始后，在有些多肽中被切除。

forword mutation　正向突变：由野生型等位基因突变为突变型等位基因。

frame-shift mutation　移码突变：由于碱基对的缺失或插入导致了翻译读框的改变，从而产生完全不同的产物。

fraternal (chizygotic) twin　异卵双生子：由两个卵同时受精产生的双生子，他们的遗传背景是不同的。

frequency distribution　频率分布：在群体中各种表型个体所占的比例。

fruiting body　子实体(孢子果)：真菌的一种器官，在其中发生减数分裂和产生有性孢子。

gain-of-function dominant　功能获得显性：由于基因的改变或表达模式的改变而引起的一种显性突变，不是简单的剔除隐性基因的活性。

gemate　配子：由减数分裂所产生的成熟的生殖细胞，每种都是单倍体，融合后形成二倍体合子。

gemetophte　配子体：在植物的生活周期中有性单倍体阶段通过有丝分裂可产生配子。

G-band (Giemsa-band)　G 带(吉姆萨显带)：一种分带染色技术，将中期染色体片段处理后用 Giemsa 染料染色，结果显示一定的带纹，可鉴定染色体。

GC element　GC 元件：真核生物启动子的元件，含有 5′－GGGCGGG－3′保守序列，位于转录起始点的上游。GC 元件的存在帮助 RNA 聚合酶结合在转录起始点的附近。

gene　基因：遗传的基本物质和功能单位，它携带了逐代传递的信息，它是一个转录区域和转录调节区的一段 DNA 序列。

gene amplication　基因扩增：在体细胞中染色体片段的拷贝数增加的过程。

gene bank　基因文库：来自单个基因组的 DNA 片段克隆的总和。

gene conversion　基因转变：引起基因改变的减数分裂的过程。在此过程中一个等位基因变成了同一座位的另一种等位基因。

gene dose　基因剂量：在基因组中存在的某特定基因的拷贝数。

gene family　基因家族：在基因组中来自祖先基因的全部后代基因。

gene flow　基因流：当生物从一个群体迁移到另一群体并繁殖后代，对受体群体的基因库贡献它的基因时，基因从一个群体流向了另一群体称基因流。

gene frequence (allele frequence)　基因频率：在孟德尔系群体中，一种等位基因占同一座位的所有等位基因的比例。

gene fusion　基因融合：在易位时，两个基因的 DNA 可能偶然连接在一起。基因融合产生杂种蛋白，或通过别的顺式元件(增强子)错误地调节一个基因的转录。

gene interaction　基因相互作用：几种不同的基因共同作用产生了一种表型特点。

gene locus　基因座位：在染色体上基因所占据的特殊位置。

gene marker　基因标记：遗传标记。

gene mutation　基因突变：遗传物质发生了可遗传的改变，使一个等位基因变成了另一个等位基因。

gene pool　基因库：在一个有性生殖群体中，所有个体所含的总的遗传信息。

gene redundancy　基因冗余：在 *E. coli* 中的染色体中有两到多倍的 tRNA 基因的现象。

generalized transduction　普遍性转导：一种类型的转导，能将供体的任何基因转导给受体菌。

genetic code　遗传密码：多肽中氨基酸序列所对应的 DNA 中碱基的信息。

genetic counseling　遗传咨询：对双亲可能生育患遗传病的子女的估计和解释。典型的遗传咨询是对一对夫妇的孩子发生某些特殊情况的概率作出估计。

gentic engineering　基因工程：通过将一个个体的基因进行修饰、插入、剔除、封闭和重组等改变细胞或个体的基因结构。在某些情况下外源基因的连接可产生新的基因重组。

genetic map (linkage map)　遗传学图、基因图谱：在连锁结构中描述非等位基因座位之间的遗传距离。

genetic mapping　遗传作图：利用遗传杂交来确定染色体上两个连锁基因之间的关系。

genetic marker (gene marker)　遗传标记：以某个等位基因为探针检测一个个体、一个组织、一个细胞、一个核或一条染色体是否存在其等位基因。

genetic recombination　基因重组：两个不同基因型的亲本产生的后代中基因产生重组的过程，如亲

本 *AB* 和 *ab* 产生 *Ab* 和 *aB* 的后代。

genetic variance (V_G) 遗传方差：在群体中不同基因型的存在所导致的表型方差。

genetics 遗传学：研究基因遗传和变异的科学。

genome 基因组：在一套染色体中遗传物质的总和。

genomic library 基因组文库：含有基因组各 DNA 片段的分子克隆的总和。

genotype 基因型：一个生物的遗传结构。

genotype frequence 基因型频率：在群体中不同基因型所占的比例。

genotypic sex determinant 基因型性别决定：在遗传和性别决定中，性染色体发挥决定性作用。

germinal mutation 生殖细胞突变：突变有的发生在体细胞，有的发生在生殖细胞中，后者称生殖细胞突变。只有这种突变可遗传。

Goldberg-Hogness box (TATA box) TATA 框：Goldberg 和 Hogness 发现的真核启动子序列，位于转录起始位点 -30 左右，保守序列为 TATAAAAA。

Group 1-intron self-splicing 第一类自我剪接内含子：mtDNA、叶绿体 DNA、四膜虫和绒泡菌 rRNA 等 RNA 的内含子可自我剪接，边界序列为 5′↓GU…AG↓，有中部核心结构和引导序列的存在，但要有 GTP 的存在。

gynandrommorph 雌雄嵌合体：一种雌性和雄性结构嵌合的个体。这些病例常是性染色体的嵌合体，如有些细胞含有雌性的染色体，有些细胞含有雄性的染色体。

half-chromosome conversion 半染色体转变：一种类型的基因转变，推测由于在真菌中存在不相同姊妹孢子，表现非孟德尔等位比例。

haploid 单倍体：具有一个拷贝的各染色体的细胞和个体。

Hardy-Weinberg law 哈迪－温伯格定律：是孟德尔定律的延续，由 Hardy 和 Weinberg (1908, 1909) 提出在一个不发生突变、迁移和选择的无限大的相互交配的群体中，基因频率和基因型频率将逐代保持不变。

harlequin chromosome 花斑染色体：在 DNA 复制时加入 5－溴尿嘧啶 (5－Brdu)，则 5－Brdu 可插入 DNA 分子中。当双链都插入 5－Brdu 时则染不上吉姆萨，而一条插入 5－Brdu 则可染色。若细胞的培养基加入 5－Brdu 则第二复制周期时，两条姊妹染色体染色状况不同，一明一暗，称花斑染色体。

helix-loop-helix (HLH) protein 螺旋－环－螺旋蛋白：一种可和真核生物 DNA 结合的转录调控蛋白。蛋白以二聚体存在，每个蛋白都有两个 α 螺旋，其间有一个环区，两个蛋白的螺旋内侧为疏水区，彼此结合。

helix-turn-helix (HTH) protein 螺旋－转角－螺旋蛋白：一种和 DNA 结合的调控蛋白。由串联的 α 螺旋构成，α 螺旋之间形成转角，HTH 常结合于大沟，常为阻遏蛋白。

hemizygous gene 半合子基因：在雄体中的 X 连锁基因，因只有一条 X 染色体，所以没有与之互补的等位基因，故称半合子基因。

hemoglobin (Hb) 血红蛋白：大部分动物的血液中传递氧和二氧化碳的珠蛋白。

hemophilia 血友病：人类中的一种隐性 X 连锁遗传病，表现为血液凝固机制缺陷。

heritability 遗传率、遗传力：在群体中遗传因子对表型方差的影响程度，或占有的比例。

hermaphroditic 两性的：动物中同时具有睾丸和卵巢的个体。在植物中同一朵花中具有雌蕊和雄蕊的物种。

hetetochromatin 异染色质：在细胞周期中保持凝聚状态的染色质，通常无活性。

heteroduplex 异源双链：由不同来源的两条 DNA 单链退火而形成的 DNA 双螺旋。若两条链之间结构不同会出现环或泡。

heteroduplex DNA model 异源双链 DNA 模型：一种解释交换和基因转变的模型，模型假设在交叉的附近异源双链 DNA(由两亲本 DNA 形成)有短的延伸产生了交换和基因转变。

heterogametic sex 异配性别：具有两种不同型性染色体的性别(如 XY 雄性)，此个体可产生带有不同性染色体的配子。

heterogeneous nuclear RNA (hnRNA)　不均一核 RNA：在核内存在的一大群不同大小的 RNA 分子，包括 mRNA 前体和其他类型的 RNA。

heterokaryon　异核体：一种培养细胞，在胞质中含有两个不同的核。

heteromorphic chromosome　异型染色体：一对大小、形态或染色不同的同源染色体。

heterplasmon　异质体：一个细胞含有遗传性不同的胞质、线粒体或叶绿体。

heterosis　杂种优势：杂合体基因型在生长、生育能力、可育性等几种性状都优于纯合体的现象。

heterothallic fungus　异宗配合真菌：有的真菌有两种交配型，有性生殖必须在两种不同交配型之间进行。

heterozygous　杂合的：二倍体生物在一个或多个座位上有不同的等位基因。

heterozygosity　杂合性：在一个座位上杂合体的比例。

Hfr (high-frequency recombination) cell　高频重组细胞：*E. coli* 中一种雄性的细胞，其 *F* 因子已整合到细菌的染色体中。当和 F^- 细胞结合时，供体基因以较高频率传递到受体细胞中，并进行重组，产生接合后体。

highly repetitive sequence　高度重复序列：在真核生物中存在的一类重复序列，重复次数为 $10^5 \sim 10^7$ 之间。

histocompatibility antigen　组织协同抗原：决定组织移植是接受还是排斥的一种抗原。

histone　组蛋白：一种碱性蛋白，真核染色体中它和 DNA 形成核小体。

homeobox　同源异型盒：在调节发育的蛋白编码序列中存在的 180bp 的保守序列。

homeodomain　同源异型区域：在很多转录因子中存在的高度保守的 60 个氨基酸序列，这些转录因子能形成 HTH 结构与特异序列 DNA 结合。

homeotic mutations　同源异型突变：改变体节特征的突变，使一个体节变成另一体节的拷贝。

homogametic sex　同配性别：带有一对同源性染色体的性别，如 XY 型中的雌性 (XX) 和 ZW 型的雄性 (ZZ)。

homolog　同系物：一对同源染色体中的每个成员。

homologous chromosome　同源染色体：在减数分裂中相互配对的染色体，或在进化中来源同一祖先，含有的基因座位完全相同的染色体。

homozygote　纯合子：具有纯合基因的个体。

hormore response element (HRE)　激素应答元件：激素结合于受体可作用转录因子，HRE 是供激素受体复合体结合，顺式调节的 DNA 序列。

host range　宿主范围：特定品系的噬菌体只能感染一定品系的细菌。

human genome project (HGP)　人类基因组 (测序) 计划：是 Watson 提出人类基因组全序列分析的计划。人类基因组约有 33×10^9 bp，含有 5 ~ 10 万的基因。

Huntington's disease　亨廷顿舞蹈病：人类常染色体显性遗传病，初期为面部运动和手势增加，逐渐发展为不随意舞蹈动作，晚期出现吞咽困难和痴呆，五十岁左右死亡。

H-y antigen　H-Y 抗原：Y 染色体连锁的 H-Y 座位 (histocompatibility locus) 的产物。20 世纪 90 年代前被认为是睾丸决定因子的候选者，现被否定。

hybrid　杂种：(1)杂合子；(2)任何不同基因型的亲代杂交所产生的个体。

hybrid dysgenesis　杂种不育：某些含有 P 因子的果蝇和正常果蝇杂交会导致后代不育，染色体畸变等缺陷，称杂种不育。

hybridization *in situ*　原位分子杂交：用同位素标记的特定基因和中期染色体杂交来确定此基因在染色体的位置。

hydrogen bond　氢键：一种弱键，存在于一个与氧或氮原子的共价键连接的氢原子和一个带负电原子之间。

hydroxyapatite　羟基磷灰石：能和 dsDNA 结合的一种磷酸钙。

hypermorph　超效等位基因：一种突变基因，效应与野生型基因相同，但表达量远远高于正常基因的产物。

hyperploid 超体：含有少量额外染色体的非整态体

hypersensitive sites (hypersensive regions) 超敏感位点：具有转录活性的基因周围的 DNA 区域对 DNase I 降解高度敏感。

hypervariable region (HVR) 高变区：在免疫球蛋白的可变区中的部分位点，它们决定抗体的特异性。

hypomorph 亚效等位基因：比正常基因产量减少的突变。

hypoploid 亚倍体：丢失了少量染色体的非整倍体。

identical (monozygotic) twins 同卵双生子：由一个受精卵经第一次卵裂产生两个单独的细胞，并发育成正常个体。同卵双生子的遗传背景完全相同。

imaginal discs 器官芽、成虫盘：在果蝇的囊胚中，一些未分化的细胞将发育为成体的组织和器官。

immunoglobulins (IG) 免疫球蛋白：通过 B 细胞分泌的一种特殊的蛋白(抗体)，存在于血液和淋巴系统中，它们负责体液免疫。

inbreeding 近交：亲缘关系很近的个体之间交配。

inbreeding coefficient 近交系数：同一个祖先的一个基因拷贝在某后代中得到纯合的概率。

incomplete dominance 不完全显性：杂合子 (Aa) 的表型显示出介于两种纯合 (AA 和 aa) 的表型之间的状况。

independent assortment 自由组合(孟德尔第二定律)，即不连锁的非等位基因在形成配子时自由组合。

induced mutation 诱发突变：由于诱变剂的处理而产生的突变。

inducer 诱导物：一种环境因子能触发操纵子的转录。

induction 诱导：基因产物是否作为对诱导物的反应。

insertion sequence (IS) 插入序列：在原核生物中的一种最简单的转座因子，它插入基因后使基因失活，它不带有抗性基因。

interference 干涉：染色体发生的第一次交换抑制了邻近域区的第二次交换。

intergenic suppressor 基因间抑制：一种基因突变效应抑制另一个基因突变的表型后果，而这两个基因的位置是不同的。

internal control region (ICR) 内部控制区：RNA 聚合酶Ⅲ识别的启动子序列，它位于基因的内部，如真核生物的 tRNA 基因和 5S rRNA 基因。

interrupted mating 中段杂交：通过搅拌中段细菌的杂交以确定不同时间内供体中基因进入受体的概率来对染色体作图。

intervening sequence (Ivs) 间插序列：即内含子，位于基因内部，不编码基因产物，在成熟 mRNA 中被切除。

intrachromosomal recombination 染色体内重组：两对基因之间的杂交形成的重组，即姊妹染色体之间的重组。

intron 内含子：见 intervening sequence 。

inversion 倒位：一种和染色体结构有关的突变，即染色体上某片段在原来位置上发生 180°翻转。

inverted repeat (IR) sequence 反向重复序列：序列相同但方向相反的序列，常存在于转座因子的两端。

isoaccepting tRNA 同工 tRNA：携带同一种特定氨基酸的各种 tRNA 分子。

isochromosome 等臂染色体：一种畸变产生的染色体，两臂的长度和结构完全相同。

isogenic strain 纯合品系：一组基因型相同的个体。

karyotype 核型、染色体组型：一个个体或细胞具有的独特的染色体的形态和数目，如在有丝分裂中期所见的情况。

kinetochore 着丝粒：见 centromere。

Klinefelter syndrome 克兰费尔特综合征：精细管发育不全或先天睾丸发育不全，人类男性的一种染色体病，常由于母体在减数分裂时 X 染色体不分离而形成 XXY 合子所致。

Lagging strand 后滞链：在 DNA 复制中，合成方向与复制叉移动方向相反的链。

λdgal 带有半乳糖基因的缺陷 λ 噬菌体：λ 噬菌体带有细菌的 *gal* 基因而本身失去了某些功能。

lampbrush chromosome　灯刷染色体：两栖动物卵母细胞中的一种大型染色体，它带有很多 DNA 侧环，在显微镜下可观察到像灯刷一样。

lawn　菌苔：在琼脂平板的表面长满了一层细菌。

leader sequence　前导序列：mRNA 5′端不编码的序列。

leading strand　前导链：在 DNA 复制时，合成方向与复制叉移动方向相同的链。

leaky mutant　渗漏突变体：突变体保留部分功能，实质上是产生了亚效突变基因。

lethal gene　致死基因：一基因表达的结果导致个体死亡。

ligand-receptor interaction　配体－受体相互作用：细胞外的分子和细胞表面上的受体或靶细胞中的受体相互作用，这种作用的一种类型是在固醇类激素与它的细胞质受体或核内受体的相互作用，另一类可能是分泌的多肽配体和跨膜受体的作用。

ligase　连接酶：一种能连接核酸中段裂的磷酸二酯键的酶。

light repair　光修复：见 photo-reactivation。

line　品系、株系：一组纯系繁殖的二倍体或多倍体生物，和同一物种中其他个体的区别是某些单一表型或基因型。

LINE(long interspersed repeated sequences)　长分散重复序列：人类中中等重复序列的分散家族，长几千个 bp，在基因组中多于 2 万个拷贝。

liner retrad　直列四分孢子、直列四分子：减数分裂减数后核分裂形成的四分体，其姐妹产物相互连接。

linkage　连锁：不同的基因位于同一染色体上。

linkage group　连锁群：一组不能进行自由组合的线性排列的基因群。

locus (loci)　基因座、座位：基因在遗传图上的位置，基因在染色体上的位置。

Lyonization　莱昂作用、莱昂化：在哺乳动物中存在的一种性染色体剂量补偿机制，即雌体中有一条 X 染色体凝聚失活，使此染色体在个体发育中不起任何作用。此现象由 Mary Lyon 发现而命名。

lysogenic bacterium　溶源细菌：含有以原噬菌体状态存在的温和噬菌体的细菌，此细菌对噬菌体来说是溶源的。直到被诱导，噬菌体开始繁殖，细菌被裂解。

lysogenic pathway　溶源化途径：对于温和噬菌体来说其生活周期有两种，一种是溶源，一种是裂解。两种途径在不同的条件下交替进行。溶源途径染色体不能自主复制，而是整合在宿主染色体的特定位置，与宿主染色体同步复制。

lysogeny　溶源性、溶源状态：温和噬菌体的染色体插入宿主染色体中，噬菌体以原噬菌体状态存在。

lytic cycle　裂解周期：噬菌体的染色体从宿主的染色体上切除下来，独立复制、生长、合成各种蛋白，最后包装成很多子噬菌体，并使宿主细胞裂解。

macromolecular　大分子：大的多聚物，如 DNA、蛋白和多糖，它们的分子量至少几千 Da。

Manx　曼克司表型：杂合体脊椎发育不良，如猫无尾巴的性状，纯合致死，为常染色体显性。

mapping functions　作图函数、定位函数：一种数学公式表示连锁图中的距离和重组频率之间的关系。

map unit (mu)　图距单位：用来计算遗传图上的两个座位之间距离的单位。两个基因之间的交换频率为 1% 就是一个图距单位。

maternal effect　母体效应：母体组织的环境对后代表型的影响。

maternal influence　母体影响：母体的基因型决定了后代的表型。

maternal inheritance　母体遗传：母体的表现决定了所有后代的表型。

mating types　交配型、接合型：在低等真核生物如真菌中相当于高等生物的性别系统。不同的交配型在特殊座位上 (MAT) 带有不同的等位基因，只有在不同的交配型之间才能交配或接合。

Maxam－Gilbert sequencing　化学测序法：由 Allan Maxam 和 Walter Gilbert 提出，他们采用不同的化学试剂使已标记的 DNA 中特异核苷酸的糖苷键断裂，然后水解使 AP 位点的磷酸二酯键断裂，经电泳和放射自显影读出 DNA 的序列。

mean 平均数、平均值：在数量遗传中用来计算方差的一个参数，以 $\overline{X}$ 表示 $\overline{X} = (X_1 + X_2 + \cdots X_n)/n$

meiosis 减数分裂：在真核生物中由二倍体的核经分裂产生单倍体配子的过程。

meiosis I 第一次减数分裂：包含前期Ⅰ、中期Ⅰ、后期Ⅰ和末期Ⅰ，在此过程中染色体数减半。

meiosis II 第二次减数分裂：第一次减数分裂后，每个子细胞中每条已复制的染色体着丝粒再分开，使姊妹染色单体彼此分离，趋向两极，最终产生四个子细胞的过程，其实质是有丝分裂。

Mendelian ratio 孟德尔比率：符合孟德尔定律的后代表型的比率。

Mendelian population 孟德尔群体：一个相互交配的群体，它们有共同的基因库。孟德尔群体是群体遗传的研究单位。

merogenete 部分二倍体：具有全部的内基因子和部分外基因子的细菌细胞，也称部分合子。

merozygote 部分合子：同部分二倍体。

messenage RNA (mRNA) 信使 RNA：含有编码信息的 mRNA，是蛋白翻译的模板。

metacentric chromosome 中着丝粒染色体：着丝粒位于染色体的中部，两臂长度相近。

metameres 体节：在较高等的动物(如果蝇)身体上可分的重复单位。

metephase 中期：在有丝分裂或减数分裂中染色体排在赤道板上的阶段。

methlation 甲基化：分子加上甲基而被修饰的过程。

migration 迁移：生物从一处到另一处。

minimal medium 基本培养基：只含有野生型生长繁殖所需成分的培养基。

missense mutation 错义突变：一种突变使某个密码子改变成编码另一种氨基酸的密码子。

mitochondrial cytopathies 线粒体细胞病：由于线粒体 DNA 发生点突变或缺失引起的人类母体遗传疾病。

mitochondrion 线粒体：真核生物细胞器，是 ATP 合成及柠檬酸循环进行的场所。

mitosis 有丝分裂：在体细胞中进行的核分裂，产生的子细胞和亲代细胞完全相同。

mitotic crossover 有丝分裂交换：在二倍体的有丝分裂中，同源染色体配对产生重组。

mobile genetic element 可移动的遗传因子：见转座因子。

moderately repetitive sequence 中等重复序列：在基因组中具有 $10^2 \sim 10^5$ 拷贝的重复序列。

molecular cloning 分子克隆：见克隆(cloning)。

molecular genetics 分子遗传学：在分子水平研究基因的结构和功能。

monocistronic mRNA 单顺反子 mRNA：真核生物的 mRNA，一个 mRNA 只编码一个蛋白。

monohybrid cross 单基因杂种杂交：亲本为单个座位杂合子的杂交，即 Aa × Aa。

monoploid 一倍体：只有一套染色体的细胞或个体。

monosome 单体：二倍体个体或细胞丢失某条染色体。

mosaic 嵌合体：一个个体不同组织中核型不同，常由染色体的不分裂或丢失造成。

mt DNA 线粒体 DNA：线粒体中一种环状 DNA。

mu phage *mu* 噬菌体：一种环状 DNA 噬菌体，可插入宿主的多个位点，并引起转座、失活和染色体重排。

multigene family 多基因家族：在进化过程中，由一个祖先基因组经重复而产生的一系列相关的基因。

multiple alleles 复等位基因：一个基因突变为 2 种以上不同的等位基因形式。

multiple-factor hypothesis 多因子学说：一种解释数量变异的学说，认为数量变异是由于大量基因相互作用，而每个基因带有很小的相加效应。

mutagen 诱变剂：一种能增加突变率的试剂。

mutant 突变体：发生突变的细胞或个体。

mutant allele 突变等位基因：任何和野生型基因不同的基因。

mutation 突变：除重组外任何使遗传物质发生改变的过程。

mutation frequency 突变频率：群体中突变的频率。

mutation rate 突变率：在每个单位时间(如每个世代)细胞发生突变的数量。

muton　突变子：最小的突变单位，现已知为一对碱基。

narrow heritability　狭义遗传力：加性方差占表型方差的比例。

natural selection　自然选择：基因型的不同繁殖。

negative assortative mating　选异交配：表型不同的亲本优先交配。

negative control　负调控：由终止或关闭转录的因子介导的调控。

neomorph　新效等位基因：经突变基因产生新的亚基或结构。

neurofibromatosis　神经纤维病：人类皮肤上一种神经细胞肿瘤和浅褐色斑，是生殖细胞中等位基因突变而产生的，是一种常染色体显性遗传。

neutral mutation　中性突变：基因序列中密码子的改变并没有改变产物的功能，即不影响生物适应性的突变和不影响生物表型效应的突变。

neutral mutation hypothesis　中性突变学说：由日本的木村资生提出的一种解释多态平衡的学说，以取代经典模型，认为蛋白存在广泛的遗传变异，但变异对于自然选择而言是中性的。

nick translation　缺刻平移：用 DNA 聚合酶 I 制备放射性标记 DNA 双链的过程，是制备放射性探针的一种方法。

noncontributing alleles　非贡献等位基因：对数量性状不起任何作用的等位基因。

nondisjuction　不分离：同源染色体或姊妹染色单体在后期不分离。

nonlinear tetrad　非线性四分子：减数分裂产生的四分子特殊序列排列。

non-Mondelian inheritance　非孟德尔遗传：核外染色体的遗传，其遗传方式不同于核遗传。

nonparental-ditype (NPD)　非亲二型：含有两种基因型的四分体，而这两种基因型却是重组型。

nnsense codon　无义密码子：链的终止密码。

nonsense mutation　无义突变：基因发生突变，使编码区中的某个密码子变为无义密码子。

nontranscribed spacer (NTS) sequence　非转录间隔区序列：在 rDNA 中转录单位之间的非转录序列，NTS 中的序列对 rDNA 转录的控制是很重要的。

Northern blot analysis　Northern 杂交：和 Southern blotting 相似的技术，不同的是用来分析 RNA 而不是 DNA，将电泳的 RNA 转移到膜上，供探针杂交。

nu body　钮体：见核小体 (nucleosome)。

nuclease　核酸酶：一种通过切断磷酸二酯键降解 DNA 的酶。

nucleofilament　核丝：直径为 30 nm 的染色质丝，由 DNA 绕在核小体核心的周围而形成。

nucleoid　类核：细菌细胞质中的 DNA 团，由支架蛋白和四周的 100 个 DNA 环构成。

nucleoplasm　核质、核浆：细胞核中半固体基质。

nucleolar organizer　核仁组织者、核仁形成区：染色体上的一个区域，和核仁相连，含 rRNA 基因。

nucleolus　核仁：在核中的细胞器，含 rRNA 及已扩增的多拷贝 rRNA 基因。

nucleoside　核苷：DNA 中的成分，由戊糖和碱基构成。

nucleosome　核小体：真核生物染色体的基本结构单位，由八聚体组蛋白和盘绕在组蛋白上的 DNA（共 200bp）及连接 DNA 构成。

nucleosome core particle　核小体核心颗粒：146 bp DNA 片段与八聚体组蛋白（$2\times H_2A-H_2B+H_3-H_4$）结合的复合体。

null allele　无效等位基因：一个等位基因不能合成产物或产物无功能。

nullisomy　缺体、零体：一种非整倍体的畸变，个体缺失了某条同源染色体。

ocher codon　赭石密码子：无义密码子 UAA。

Okazaki fragment　冈崎片段：在 DNA 不连续复制中，后随链上先合成的短的 DNA 片段，然后由这些片段连接成连续 DNA 链。

oligomers　寡聚体：短的 DNA 分子。

oncogene　癌基因：细胞中某些具有正常功能的基因，但突变后会导致细胞癌变。

one gen one enzyme hyoithesis　一基因一酶学说：Beadle 和 Tatum 根据红色链孢霉的研究提出的，认为基因控制性状的本质是每个基因控制了一种酶的合成。以后发现酶可由多肽构成，故改为

一基因一多肽学说。

opal codon 乳白密码子：无义密码 UGA。

open reading frame (ORF) 可读框或开放读框：在 DNA 片段中一种潜在的蛋白编码序列，它具有起始密码子、终止密码子和二者之间的密码子。

operator 操纵基因：在操纵子中、与启动子相连的调节位点，可供反式调节蛋白（如阻遏蛋白）识别结合来调节原核基因的转录。

orchered tetrads 顺序四分子：在链孢霉的子囊中减数分裂的产物——子囊孢子按一定顺序排列。

origin of replication 复制起点：原核生物的染色体上的特异区域，它是 DNA 复制蛋白的识别和结合区域，复制也是从此区域开始进行的。

outbreeding 远交：亲缘关系很远或无关的个体之间的杂交。

overdominance 超显性：杂合子的适应性超过两种纯合子的现象。

ovum 卵、卵子：成熟的卵细胞。

P element P 因子：果蝇的转座因子，可用于插入诱变和配子转化的工具。

P generation 亲代：杂种 1 代的双亲。

pachynema 粗线期：减数分裂前期中的一个阶段，同源染色体配对交换。

paracentric inversion 臂内倒位：倒位片段在一条染色体上、倒位片段不含着丝粒。

parental ditype (PD) 亲二型：链孢霉等真菌的四分体出现两种基因型，和双亲的基因型相同。

partial dominance 部分显性：即不完全显性。

partial diploid 部分二倍体：见 merozygote。

paticulate factors 颗粒因子：即基因，开始时孟德尔假设的遗传因子，既是结构单位又是突变单位和重组单位，而现代基因仅是功能单位。

pedigree analysis 谱系分析：根据显示出个体的不同表型的家族谱系来分析遗传特征。

penetrance 外显率：在一个群体中带有特殊基因型的个体显示出相应表型的比例。

peptide bond 肽键：连接两个氨基酸的键。

peptidyl transferase 肽转移酶：在蛋白合成中能催化形成肽键的酶。

pericentri inversion 臂内倒位：一个倒位片段涉及到染色体的两个臂，含着丝粒。

permissive conditions 允许条件：在条件突变下显示出野生型表型的环境条件。

petite colony 小菌落：酵母中的一种突变型，线粒体 DNA 突变产生的是细胞质小菌落，核 DNA 突变产生分离型小菌落（segregational petites）。

phenocopy 拟表型、表型模拟：环境导致的一种表型变化和某种由突变型产生的表型相似。

phenotype 表现型：由基因型产生的外在的遗传性状。

phenotypic sex determination 表型性决定：环境对某种生物性别决定起重要作用的过程。

phenotypic variance (V_P) 表型方差：对一个性状观察到的总方差。

phenylketomuria (PKU) 苯丙酮尿病：由于编码苯丙氨酸羟化酶基因突变所引起的人类代谢病，为常染色体隐性遗传。

Philadelphia chromosome 费城染色体：人类 9 号和 22 号染色体长臂之间的移位。常出现在慢性髓细胞性白血病病人的血细胞。

phosphodiester bond 磷酸二酯键：DNA 或 RNA 中糖基和磷酸基之间的键，这种键形成了糖磷酸骨架。

photoreactivation (light repair) 光修复：一种修复胸苷二聚体的方法，波长位 320～370nm 的可见光可直接使胸苷二聚体恢复。

piebald 花斑（无色）：哺乳动物皮毛的一种表型，由于缺乏黑色素而出现的白斑，为常染色体显性遗传。

pilus 性伞毛：F 因子使细菌表面长出 2～3 根性伞毛，长 2～3μm、外径 8nm、内径 2nm，用来和 F－细菌接合，或被单链 DNA 噬菌体吸附，但不能为 DNA 的转移提供通道。DNA 的转移可能通过膜上的蛋白形成的孔道。

plaque　噬菌斑：由于噬菌体裂解菌苔而形成的圆形区域。

plasmid　质粒：核外的遗传因子，由双链 DNA 组成，可自我复制。

plate　平板：(1)用来培养微生物的平皿；(2)在平板表面的固体培养基上涂布细胞。

pleiotropic mutation　多效突变：使不同性状发生改变的突变。

point mutation　点突变：由于一个碱基的改变导致的突变。

poky mutation　生长缓慢突变株：在脉孢菌中生长缓慢的线粒体突变。

polar mutation　极化突变：在细菌的操纵子中上游的基因发生无义突变、移码突变或插入突变不仅影响该基因的活性，也使下游基因的产物减少或改变。

polyA polymerase　聚腺苷酸聚合酶：催化产生 mRNA 3′端多聚腺苷的酶。

polyA tail　多聚腺苷尾巴：真核生物的 mRNA 转录后在 3′端加上的多聚腺苷序列。

polycistronic mRNA　多顺反子 mRNA：一条 mRNA 编码多个蛋白。

polydactyly　多指(趾)：人类的一种显性突变，表现为多于五个指(趾)的表型。

Polygene　多基因：在数量遗传中，一种学说认为可能由多个微效基因在控制数量性状。

polylinker　多位点接头：在克隆载体上接上一段带有很多限制性酶切位点的序列，供插入外源 DNA 用。

polymerase chain reaction (PCR)　多聚酶链式反应：一种扩增 DNA 的方法，可在混合的 DNA 中特异地扩增靶序列。

polymorphism　多态性：在同一个互相交配的群体中，存在两种或更多的基因型。

polypeptide　多肽：一条氨基酸连成的链，即一个蛋白。

polyploid　多倍体：有三套或三套以上染色体的细胞或个体。

polyribosome (ploysome)　多核糖体(多体)：蛋白合成时与 mRNA 相结合的一串核糖体。

polytene chromesome　多线染色体：在一些特殊的细胞中(如果蝇和摇蚊的唾腺中)，通过核内有丝分裂，DNA 大量复制，但胞质不分裂，染色体实质由多条染色体排在一起而产生。

population genetics　群体遗传学：遗传学的一个分支，以数学方式来讨论群体孟德尔式遗传的结果。

position effect　位置效应：一个基因随染色体畸变而改变它与邻近基因的位置关系，从而改变表型效应的现象。

position effect variegation　位置效应花斑：在某些细胞中由于染色体畸变使某些基因移到异染色体的附近而不能表达，从而形成花斑。

positive assortative mating　正选交配：表型相似的个体之间的交配，其发生频率比随机选择的交配要高得多。

positive control　正控制、正调节：促进转录的调节。

posttranslation transport　翻译后运输：蛋白质翻译后转运到细胞器(核、线粒体、叶绿体)中或进入粗面内质网转运到液泡、溶酶体、膜或膜外，在原核生物中可分泌到胞外。

precursor RNA molecule　前体 RNA 分子：尚未加工的 RNA 初始转录本。

Pribrow box　Pribrow 框：原核启动子中 -10 序列，由 Pribrow 发现，位于转录起始点上游 10 个碱基附近，保守序列为 TATAAT。

primary nondisjunction　初级不分离：一种罕见的现象，姊妹染色体或染色体在减数分裂中不分离同趋向一极，结果产生了异常的配子。

primary transcribe　初级转录本：见前体 RNA 分子。

primer　引物：DNA 合成时，任何 DNA 聚合酶都不能从头开始，而只能在 3′-OH 上延伸 DNA，因此必须存在短片段的 RNA 或 RNA 结合在模板上为合成提供 3′-OH，这就是引物。

primosome　引发体：*E. coli* 的引发酶、解旋酶和其他一些多肽在一起构成引发体，催化 DNA 合成的起始。

probe　探针：在分子杂交中用以检测特异序列的标记 DNA 片段。

product rule　相乘法则：两个独立事件同时发生的概率等于各自概率的乘积。

proflavin　硫酸黄原素：一种诱变剂，可插入 DNA 分子中引起移码突变。

prokaryotic cell 原核细胞：细胞没有核膜，如细菌和蓝藻的细胞。

promoter 启动子：基因 5′端的转录调控区，是 RNA 聚合酶识别和结合的特异区域。

prophage 原噬菌体：整合在宿主染色体中的噬菌体，这种状态的噬菌体叫原噬菌体。

propositus (proband) 先征者：在人类遗传学中研究某一性状在一个家族首先发现的第一例有此性状者。

protooncogene 原癌基因：正常细胞中的癌基因，功能是调控细胞的正常增生和信息传递，当发生突变后就可使细胞癌变。

prototroph 原养型：即野生型，在基本培养基上可生长。

provirus 原病毒：整合在宿主细胞染色体中的病毒。

pseudodominance 拟显性：由于一个显性基因的缺失，使其相应的隐性等位基因得以表达的现象。

pseudogene 假基因：由一个祖先基因经过重复突变产生了基因家族，家族中有些成员突变，永久性失去活性，此称假基因。

pulse-chase experiment 脉冲追踪实验：在有放射性培养基中生长一个短暂时间，然后再移到非放射性培养基中生长一段长时间。

Punnett square 庞纳特方格：用棋盘方式计算 F_2 代基因型及表型比例的方法。

pure-breeding line or strain 纯种繁殖系或品系：一群遗传上相同的个体，它们互相杂交产生的后代表型相同。

Q banding Q 带：一种染色体荧光染色分析技术。

Quantitative genetics 数量遗传学：研究数量性状的遗传规律。

Quantitative (continuous) traits 数量性状：显示出连续的可量度的性状，且与环境关系密切。

R plasmid 抗药性质粒：带有转座因子的质粒，转座因子上带有抗药性标记。

random mating 随机交配：在群体中基因型之间的交配以基因型频率为比例进行。

rDNA repeat units rDNA 重复单位：18S、28S、5.8S rRNA 基因串联重复排列这些染色体上。

reading frame 读框：从特定的起始密码子开始由 3 个核苷酸一组组成的编码序列。

recessive 隐性：只有在纯合时才能表达的等位基因或表型。

recessive lethal allele 隐性致死基因：一种等位基因，纯合时可致死。

reciprocal cross 正反交、互交：两次杂交，其亲本的基因型相互颠倒一下，如 A × B，B × A。

recombinant 重组体、重组子：通过重组产生的不同于亲本基因型的个体或细胞。

recombinant DNA 重组 DNA：两个不同源的 DNA 分子通过重组产生

recombinant frequency (RF) 重组率：重组细胞或个体的比例。

recombination 重组：由于独立分配或交换在后代中产生新的基因组合。

recombinational repair 重组修复：通过重组酶类进行重组的过程来修复 DNA 的损伤。

recon 重组子、交换子：重组的最小单位，现在认为是一个核苷酸对。

redundant DNA 冗余 DNA：见重复 repetitive DNA。

regression 回归：由 Gaiton 提出的一种分析数量性状的方法，表示变量对自变量的依存关系。

regression coefficient 回归系数：两个相关变量的变化率，即回归线的斜率。

regulated gene 被调节基因：一个基因的活性因细胞或机体的需要而受到控制。

regulatory factors 调节因子：一种蛋白，可激活或抑制基因的转录。

regulatory genes 调节基因：与结构基因的打开和关闭有关的基因。

release factor 释放因子：蛋白合成时，当核糖体移动到终止密码子时，没有相应的氨酰 tRNA 进入 A 位，而一种蛋白因子可进入，促使多肽的释放和核糖体的解离，此蛋白称释放因子。

replica plating 影印培养：将一个平板上的菌落按原来的分布转移到另一平板上的方法，即将丝绒包在与平板匹配的圆柱体上，灭菌后在有菌落的平板上印一下再转印到另一平板上，丝绒上无数的纤维起到接种针的作用。

replication bubble 复制泡：在复制中相对的两个复制叉之间的已变性的 DNA 区域。

replication fork 复制叉：复制时 DNA 双链解开形成的 Y 形结构。

replication machine (replisome) 复制机制(复制体):在 DNA 复制中由两个分子的 DNA pol Ⅲ和解旋酶、引发体结合形成的复制体。

replicon 复制子:一个复制单位,包括真核的复制起点和位于两侧的复制终点之间的区域。

reporter gene 报告基因:一种表型表达易于测定的基因,用于研究重组 DNA 中启动子在不同的时间、不同发育阶段的活性。

repressor protein 阻遏蛋白:一种可结合在操纵基因上并阻止转录的分子。

repulsion 互斥:两个连锁的杂合基因对,排列方式为 Ab/aB。

restriction enzyme 限制性酶:能在特异位点切开双链 DNA 的一种酶,它对于分析 DNA 和构建 DNA 重组分子都是很重要的。

restriction condition 限制条件:一种环境条件,在此条件下突变会产生突变表型。

restriction fragment length polymorphisms (RFLP) 限制性片段长度的多态性:在群体中不同个体基因的侧翼序列不同,但可稳定遗传,若用限制性酶来切可得到很多不同长度的酶切片段,它们可作为一种标记。

restriction map 限制性酶切图:也称物理图谱,显示出基因组中的 DNA 上不同酶切位点的相互关系。

retroposon 反转录子、反转录转座子:由一个 RNA 分子经反转录产生的转座因子。

retroviruses 反转录病毒:单链 RNA 病毒。先经反转录成 ds DNA 的中间体,插入到宿主的染色体中再进行转录。

reverse genetic 反向遗传学:为研究一个基因或蛋白的正常功能,可将它们的突变形式引入细胞,观察其表达,此方法称反向遗传学。

reverse mutation (reversion) 回复突变(回变):一种突变使突变基因变成了野生型基因。

reverse transcriptase 反转录酶:一种以 RNA 为模板合成 DNA 的聚合酶。

ribonucleases (RNase) 核酸酶:催化 RNA 降解的酶。

rho ρ 因子:在 *E. coli* 中转录的弱终止子,需 ρ 因子的协助才能终止转录。

ribonucleic acid (RNA) 核糖核酸:是多聚核糖核酸分子,单链,戊糖为核糖,四种碱基与 DNA 不同的是含有 U 而无 T。

ribosomal DNA (rDNA) 核糖体 DNA:编码 rRNA 的 DNA 区域。

ribosomal protein 核糖体蛋白:与 rRNA 结合形成核糖体的蛋白。

rRNA 核糖体 RNA:一种 RNA 分子,由核仁形成区编码,对核糖体的结构和功能都有一定的作用。

ribosome 核糖体:由 rRNA 和核糖体蛋白构成的复杂的细胞器颗粒,是蛋白合成的场所。

ribosome-binding site 核糖体结合位点:mRNA 分子上的一段序列,在此处核糖体定向正确阅读读框并起始蛋白合成。

ribozyme (RNA enzyme) 核酶、核糖核酸拟酶:具有自我剪接功能的 RNA 分子。

R looping (R loops) R 环:由 M Thomas, R White 和 R Davis 发展的技术,当双链 DNA 在变性温度下孵化,部分区域双链打开,而单链互补的 RNA 分子能和相应的 DNA 单链结合,被取代的 DNA 单链形成 R 环。

RNA editing RNA 编辑:RNA 的一种加工,以引导 RNA 为模板,在 RNA 上增加或减少几个碱基的过程。

RNA ligase RNA 连接酶:内含子从前体 RNA 切除后,催化外显子部分彼此连接的酶。

RNA polymerase RNA 聚合酶:催化 RNA 合成的聚合酶,它和 DNA 聚合酶不同,可从头起始,无需引物。

RNA plymerase Ⅰ RNA 聚合酶Ⅰ:存在于真核细胞的核仁中,催化 18S、5.8S 和 28S rRNA 基因转录。

RNA plymerase Ⅱ RNA 聚合酶Ⅱ:仅存在于真核细胞的核质中,催化 mRNA 编码基因的转录。

RNA plymerase Ⅲ RNA 聚合酶Ⅲ:仅存在于真核细胞的核质中,催化 tRNA 和 5S rRNA 基因转录。

Robertsonian translocation 罗伯逊易位:一种非交互易位,两个非同源的近端着丝粒染色体的长臂连接成一条单着丝粒的中央着丝粒染色体,又称整臂易位。

sample 样本:提供群体信息的亚单位,样本要求大小合适,并随机取样才具有代表性。

sampling error 样本误差:在一个小样本中预期的比例会发生随机改变的现象。

satellite DNA 卫星 DNA:真核生物基因组中的一种高度重复序列,富含 A－T ,当进行 CsCl 密度梯度离心时,基因组呈现一条宽的带,而在其上方高度重复序列显示了单独的一条细带,故称卫星 DNA。

scaffold attachmentation region(SARs) 骨架附着区:DNA 上的特异位置,附着在染色体的骨架上。

secondary law 第二定律:见自由组合定律(independent assortment)。

secondary nondisjunction 次极不分离:初极不分离产生的雌性后代中 X 染色体再度不分离。

second－site mutation 第二位点突变:见抑制基因突变(suppressor mutation)。

selection coefficient 选择系数:计算对一种基因型的选择相对强度。

selection differential 选择差数:在自然和人工选择中,被选择亲代的表型平均值和未被选择的群体平均表型之间的差异。

self-assembly 自组装、自动装配:由亚基按特定的模式自动聚集成某种功能结构的过程。

self-fertilization (selfing) 自体受精:同一个体产生的雌性和雄性配子相互结合。

self-splicing 自我剪接:某些前体 RNA 分子中的内含子本身可进行自我切除,但此过程在有的生物中却是蛋白依赖性反应。

semiconservative replication mode 半保留复制模型:在 DNA 复制两条子 DNA 链中,每条双链都含有一条亲代的单链。

semidiscontinuous 半不连续(复制):DNA 复制时,前导链上 DNA 的合成是连续的,后随链上是不连续的,故称半不连续复制。

sense codon 有义密码子:mRNA 上相对一个氨基酸的密码子。

sex chromosome 性染色体:在真核生物中和性别相关的染色体,如 X、Y 和 Z、W。这些染色体在性别决定中起重要作用。

sex determination 性别决定:通过遗传或环境的作用使个体建立性别的过程,其实质是性腺的分化的决定。

sex-influenced traits 从性性状:一种性相关遗传,基因位于常染色体上,但基因的表达和性激素有关,因此在不同性别中基因型相同,但表型不同。

sex-limited trait 限性性状:一种性相关遗传,基因位于常染色体或性染色体上,但性状仅在一种性别中表达,这些性状常和第二性征有关。

sex-linked 性连锁:基因位于 X 和 Y 染色体非同源区的遗传。

sex reversed 性反转:在哺乳动物中的一种综合征,XX 的个体发育为雄性或 XY 的个体发育为雌性,此与 *SRY* 区域的突变有关,可能和其他基因也有关。

sexduction 性导:由于 F′因子通过接合将特定的基因以很高的频率传递给 F^- 受体的过程。

sexual reproduction 有性繁殖:涉及到由减数分裂产生的两种配子融合的繁殖。

shotgun technique 鸟枪法:将基因组 DNA 切成很多小的片段装到载体上再转化宿主细胞进行克隆,然后用特异探针通过分子杂交钓出相应的基因供研究。

shuttle vector 穿梭载体:一种克隆载体,可在两种或多种宿主中克隆,常用于重组 DNA,在 *E.coli* 中克隆,再到另一种生物中(如酵母)表达。

sickle－cell anemia 镰刀形细胞贫血:一种潜在的人类致死遗传病,是由于 β－珠蛋白的基因突变使在缺氧时红细胞变为镰刀形,为常染色体隐性遗传。

signal hypothesis 信号学说:有关蛋白通过特殊的疏水氨基酸区域跨膜分泌的学说,此疏水氨基酸区域在 ER 中被切除和降解。

signal peptidase 信号肽酶:在 ER 中切除和降解信号肽的一种酶。

signal recognition particle (SRP) 信号识别颗粒:在真核生物细胞质中一种小分子 RNA 和六种蛋白的复合体,此复合体能识别核糖体上新生肽末端的信号序列并与之结合,使肽合成停止,同时它又可和 ER 膜上的停泊蛋白识别和结合,从而将 mRNA 上的核糖体带到膜上。

signal sequence 信号序列、信号肽:分泌蛋白的 N 端存在的 20～30 氨基酸的疏水区域,提供跨膜信

号，当进入 ER 中被切除和降解。

silencer　沉默子、沉默基因：见沉默元件(silencer element)。一种转录调控元件，特点很像增强子，但不增强转录，而是减弱转录，故也称负增强子。

silent mutation　沉默突变：即同义突变，突变虽然替换了碱基，但氨基酸序列未变，保持野生型的功能。

simple telomeric sequences　简单的端粒序列：在真核生物染色体 DNA 的末端存在简单的串联重复序列。

SINES (short interspersed repeated sequences)　短的分散重复序列：在真核基因组中存在短而分散高度重复序列，长 500bp 以下，基因组中有数千个拷贝，如 *Alu* 序列。

single-strand DNA binding (SSB) protein　单链结合蛋白：一种六聚体蛋白，可与 DNA 单链结合使其不会复性。在复制过程中此蛋白起重要作用。

sister chromatid　姐妹染色单体：染色体复制后由着丝粒连接的两条姐妹染色单体。

sister-chromatid exchange (SCE)　姐妹染色单体交换：在有丝分裂或减数分裂时姐妹染色单体发生的交换，其交换频率可用来检测诱变剂。

site-specific recombination　位点特异性重组：重组的一类，只发生在特异 DNA 区域，有短的同源序列，重组的蛋白不是 rec 系统而是 int 等，如噬菌体 λ 的定点插入。

small nuclear ribonucleoprotein particles (snRNP)　核微小核糖核蛋白颗粒：由小分子核内 RNA 和蛋白构成的复合体，前体 mRNA 在上面进行加工。

small nuclear RNA(sn RNA)　核内小 RNA 、小核 RNA：真核生物核中由转录产生的四种主要 RNA 分子，它用于 mRNA 的加工。

solenoid structure　螺线管结构：真核生物染色质的二级结构，一级结构由核小体纤丝（10nm）构成，二级结构由核小体纤丝进一步螺旋形成 30nm 螺线管，每个螺旋由六个核小体组成。

somatic cell hybridization　体细胞杂交：两个不同物种的体细胞融合产生体细胞杂种供遗传分析。

somatic mutation　体细胞突变：非生殖细胞的突变，在当代会出现突变体斑点或区域，此特征不能遗传。

somatic-cell genetics　体细胞遗传学、无性遗传学：涉及研究体细胞突变、分类、交换和细胞融合。

SOS repair　应急修复、倾向差错修复：当紫外线照射引起损伤时，同时也诱导 SOS 系统，允许 DNA 可通过损伤的模板合成，但相应的 DNA 片段会产生差错，尽管如此，细胞或个体仍能存活。

Southern blot　Southern 印迹：将电泳的 DNA 片段从胶上转移到膜上供分子杂交分析。

spacer sequences　两个基因之间的 DNA，不转录，功能不明。

specialized transduction　特异性转导：只转移特定基因的转导，如 λ 噬菌体只转导 *gal* 和 *bio* 基因。

splicesome　剪接体：一系列 snRNPs 结合而成的复合体，它们结合在前体 mRNA 上，帮助核 mRNA 内含子的剪接。

splicing　剪接：切除内含子，连接外显子的反应。

spontaneous mutations　自发突变：在没有任何诱变剂的情况下而产生的突变。

spore　孢子：(1)在植物或真菌中，有性孢子是减数分裂产生的单倍体细胞；(2)在真菌中无性孢子是起配子作用或作为新单倍体个体的原始细胞的体细胞。

sporophyte　孢子体：二倍体，在植物生活周期中的无性世代，它通过减数分裂产生单倍体的孢子。

standard deviation　标准差：方差的方根。

standard error of gene frequency　基因频率的标准误：群体基因频率中变异量的计算，它是基因频率方差的方根。

steroid response element (RE)　固醇类效应元件：在基因的 5′上游固醇类激素和其受体形成的复合物所识别和结合的特异序列。

strain　株、菌株：纯合繁殖系，常用于单倍体生物，如细菌或病毒。

stop codon　终止密码子：见链终止密码子。

stringent response (stringent control)　严谨反应(应急控制)：(1)细菌在氨基酸饥饿时将大部分活性

区域关闭的反应。(2)细菌质粒拷贝控制在几个拷贝的反应。

structural gene 结构基因:编码蛋白质的氨基酸序列的基因。

submetacentric chromosome 亚中着丝粒染色体:着丝粒偏于一侧的染色体,后期是 J 形。

subvital gene 亚生活力基因:一些基因由于它的表达可引起某些个体(不是全部)的致死。

sum rule 相加法则:互斥事件的总概率等于各事件概率之和。

supercoil 超螺旋:DNA 双螺旋本身进一步盘绕称超螺旋。超螺旋有正超螺旋和负超螺旋两种,负超螺旋的存在对于转录和复制都是必要的。

superinfection 超感染:噬菌体感染一个已有原噬菌体的细胞。

supersuppressor 超抑制基因:一种突变能抑制其他多种突变、最典型的是无义突变。

suppressive peptite 抑制性小菌落突变株:一种酵母小菌落,它与野生型杂交所产生的后代具有各种非孟德尔比的小菌落,此是由于线粒体基因发生突变所致。

suppressor gene 抑制基因:一个可抑制其他基因突变的基因。

synapsis 联会:在减数分裂的偶线期阶段,同源染色体之间的特异配对。

synaptinemal complex 联会复合体:在减数分裂的前期同源染色体配对时形成的复合物。

syncytial blastoderm 合胞囊胚:在昆虫的囊胚阶段,在早期胚中四周有很多单个的核。

synsytium 合胞体:一个单个细胞,四周有很多核,核间没有隔膜把它分开。

synkaryon 融核体、合核体:随着两个不同细胞的融合而产生的核融合。

syntenic 同线的:位于同一条染色体上。

TATA box TATA 框:Goldberg - Hogness 框,是真核类启动子区中的序列,与原核类的 Pribnow 框类似,但它对转录虽是有利的,却不是必需的。

TATA factor TATA 因子:关键的转录因子(TFIID),结合在 TATA 元件上。

targeted gene knockout 定向基因的剔除:通过定点突变,在一个基因中导入无义突变,然后将这个基因导入与正常的等位基因进行同源重组并取而代之。

tautomeric shift 互变异构移位:当碱基中氮所产生的氢键的改变会导致自发的碱基错配,而引起突变,如产生 C - A、G - T 配对。

T - DNA 转移 DNA:Ti 质粒或 Ri 质粒中的一段 DNA 序列,可插入到宿主植物的基因组中。

telocrntric chromosome 端着丝粒染色体:(1)着丝粒在末端附近的染色体;(2)一条畸变的染色体。由于着丝粒的异常横向断裂使一条染色体变成了两条端着丝粒染色体。

telomere associated sequence 端粒伴随序列:在真核类染色体末端附近的区域常含有重复序列,但仍是复杂序列,延伸数千碱基对,形成染色体 DNA 的分子末端。

telophase 末期:在核分裂时,子核重新形成的最后的一个阶段。

temperate phage 温和噬菌体:能成为原噬菌体的噬菌体。

temerature - sensitive mutoation 温度敏感突变:一种条件突变,在一定的温度条件下产生突变表型,而在另一温度范围产生野生型表型。

template strand 模板链:提供 DNA 合成或转录模板的 DNA 单链,或提供翻译信息的 mRNA 链。

termination factors 终止因子:即释放因子(release factor)。

terminator 终止子、终止区:原核生物转录终止时的特殊序列。

tertrary structure of a protein 蛋白的三级结构:二级结构的折叠和超螺旋形成球状分子。

testcross 测交:将一个未知基因型的个体(常表达为显性表型)与一个纯合隐性个体杂交,以检测其基因型。

teaticular feminization syndrome 睾丸女性化:为人类性连锁隐性遗传病,是编码雄性激素受体基因发生突变所致。患者有女性外阴和第二性征,无子宫,有睾丸,位于腹股沟,阴道后侧为盲端。

tetrad 四分体、四分子:(1)在减数分裂前期和中期四条同源的染色单体;(2)减数分裂产生的四个子细胞。

tetrad analysis 四分体分析:对减数分裂所有产物的分析,可用来研究在减数分裂中染色体和基因的行为。

tetraploid　四倍体：具有四套染色体的细胞或个体。

tetratype (T)　四型：杂交后代中产生四种不同的基因型，两种为亲组合，两种为重组合。

three-point testcross　三点测交：涉及到三个连锁基因的测交，可用来测定待测基因的序列和相互之间的图距以及双交换的频率和干涉。

thymine dimer　胸苷二聚体：DNA 经紫外线照射可使相邻的两个胸苷之间共价连接形成环丁酰二聚体。

Ti plasmid　Ti 质粒：农杆菌中的一种可致瘤的环状质粒，它使细菌能感染植物细胞产生肿瘤，现已改造为一种植物转基因载体。

topoisomerases　拓扑异构酶：能切开 DNA 双链(Top Ⅱ)和单链(Top Ⅰ)的一种酶，可引入超螺旋。

totipotency　(分化)全能性：核具有通过各发育阶段产生一个正常成体的能力。

trailer sequence　非翻译区(mRNA 终止密码子后的)：mRNA 上终止密码子后的 3′端非翻译区，不同的分子长度不同。

transconjugant　转接合子：在细菌的接合过程中，供体 DNA 的受纳。

transcription　转录：遗传信息从双链 DNA 上转移到单链 RNA 分子上，也称 RNA 合成。

transcription factors (TFs)　转录因子：真核类三种 RNA 聚合酶转录起始时所必需的特殊蛋白质，每种聚合酶都需它特异的一套 TFs。

transcription terminator sequence　转录终止序列：见终止子(terminator)。

transdetermination　转决定：在发育过程中成虫盘并不全部去分化，但可转成另一种决定途径。

transduction　转导：以噬菌体为载体，基因从供体菌转移到受体菌中，并得到表达的过程。

transfection　转染：(1)分离的外源 DNA 转导入培养细胞中的过程。(2) 将裸露的噬菌体完整 DNA 导入受体菌中的过程。

transfer RNA (tRNA)　转运 RNA：小分子 RNA，可将特定的氨基酸根据 tRNA 的反密码子和核糖体 A 位 mRNA 上的密码子的互补带到核糖体中进行蛋白质合成。

transformant　转化体：通过转化产生的重组体。

transformation　转化：(1)将分离的外源 DNA 导入受体细胞并得以表达的过程。(2)正常的高等真核生物组织的培养细胞变成失去分裂控制的癌细胞样状态的过程。

transgenic organism　转基因生物：导入了外源 DNA 并产生表达的动植物。

transit peptidase　转肽酶：一种酶可从转运到细胞器的蛋白上切除转移序列。

transit sequence　转运序列：蛋白质的 N 端的额外序列，为转入到细胞器中所必需。即导肽。

transition　转换：DNA 中碱基替换的一种类型，即同类碱基(嘌呤与嘌呤之间或嘧啶与嘧啶之间)的转换。

translation　翻译：以 mRNA 为模板合成蛋白质的过程。

translocation (transposition)　易位：(1)涉及到染色体片段位置发生改变的染色体突变；(2)在多肽合成中核糖体在 mRNA 上移动一个密码子。

transmission genetics　传递遗传学：即经典遗传学，遗传学的一个分支，是研究基因逐代传递机制的学科。

transposable genetics element (TGE)　转座遗传因子：在原核和真核生物的染色体中，有一种本身可移动的遗传因子，它们可从一个地方转移到另一个地方。

transposase　转座酶：转座子 IS 因子编码的一种酶，它催化转座因子的转座活性。

transposion　转座：转座因子在染色体上的移动。

transposon　转座子：一种可移动的 DNA 片段，两侧有重复序列，中间有编码与抗药性及转座相关蛋白的基因。

transversion　颠换：在碱基替换中嘌呤替换嘧啶，嘧啶替换嘌呤。

triplet　三联体：三对核苷酸组成的一个密码子。

triploid　三倍体：一个有三套染色体的细胞，或由这种细胞构成的个体。

trisomy　三体：多了一条染色体的二倍体细胞或由这种细胞构成的个体。

trisomy－21 21－三体：见唐氏综合征（Down syndrome）。

true-breeding strain 纯系：见 pure-breeding line 或 strain。

tumor-suppressor gene 肿瘤抑制基因：一种抑制肿瘤形成的基因。如 *Rb*、*P53*、*P21* 等。

tumor viruses 肿瘤病毒：一种可导致细胞去分化并发展成肿瘤的病毒。

Turner syndrome 特纳综合征：又称先天性卵巢发育不全、染色体为 46 XO 的女性，第二性征发育不良，生长缓慢，体矮，后发际低，性腺发育不全，轻度智力障碍。

twein spots 双生斑：动物身体上紧靠在一起的两小块组织，它们的基因型不同，而且不同于周围的组织，是异合体的体细胞交换的结果。

2μ（2 micron）plasmid 2μ 质粒：在某些酵母细胞中存在着一些天然的额外环状 DNA 分子，周长为 2μ，可作为基因工程的载体。

unequal crossover 不等交换：也称非法交换，指发生在联会时略有参差的两个染色体片段之间的交换。

uniparental inheritance 单亲遗传：通常是由核外基因的传递而造成的现象，即在所有后代中只出现一种亲代的表型。

unique（single－copy）sequence 单一序列：在每个基因组中仅一个或少数几个拷贝的序列。

upstream activator sequences（UAS） 上游激活序列：在酵母和其他的一些真核生物中存在着与增强子相似的序列，它可位于启动子上游的不同距离处，功能是定向的。

URF（unidentified reading frame） 未鉴定读框：未知功能的开放读框。

variance 方差：一个变数与平均数的偏差的平方的平均值。

variance of gene frequency 基因频率的方差：在群体中一个等位基因频率的方差。

variation 变异：在一个群体中亲代与子代之间或个体之间的差异。

variegation 花斑：各种原因所引起的动植物组织部分区域表型不同。

vector 载体：在分子克隆中被用来携带外源基因的质粒或噬菌体染色体。

virulent phage 烈性噬菌体：像 T4 噬菌体一样感染细菌时总出现裂解周期的噬菌体。

VNTR（variable number tandem repeat） 可变数同向重复序列：在染色体中一种重复序列，在不同个体或同一个体中来源不同的一对同源染色体上重复的次数不同，可用来作 DNA 指纹。

wild type 野生型：在自然中或标记的实验生物中所具有的正常基因型或表型。

wobble hypothesis 摆动假说：由 Francis Crick 提出，推测反密码子的 5′端（密码子的 3′端）要求不像前两个碱基严格，可和密码子 3′端三种不同的碱基配对。

X:A ration 性指数：果蝇和线虫的性别决定与 Y 染色体无关，而是由性指数决定的，即 X 染色体的个数（X）除以常染色体的组数（A）。

X chromosome inactivation X 染色体的失活：在雌性哺乳动物的胚胎中，一条 X 染色体在早期随机失活，导致由异合的 X 连锁基因编码的功能嵌合（见剂量补偿和巴氏小体）。

X linkage X 连锁：X 染色体与 Y 染色体非同源区域上的基因的遗传方式。

X and Y linkage X 和 Y 连锁：在 X 和 Y 同源区域上的基因的遗传方式。也称不完全性连锁，同 X 和 Y 之间可能会发生交换。

X-ray crystallography X 射线（衍射）晶体分析法：通过 X 射线衍射对蛋白质或核酸晶体进行分析来研究它们的结构和功能。

Y linkage Y 连锁：在 Y 染色体与 X 染色体非同源区域上的基因的遗传方式。

Z-DNA 左旋 DNA：DNA 家族中的一种，呈左旋。

zinc finger protein 锌指蛋白：一种含有半胱氨酸的蛋白，可和锌离子结合形成指状结构，插入到 DNA 的大沟中，进行调控。

Zoo blot 动物园印迹：用一种物种的 DNA 探针通过 Southern blotting 与其他物种的 DNA 进行分子杂交，来测定是否存在同源序列的方法。

zygonema 合线期、偶线期：在减数分裂前期Ⅰ，同源染色体开始配对的阶段。

Zygote 合子：由卵和精子融合形成的细胞，为二倍体。

zygotic induction　合子诱导：当含有 λ 的 Hfr 品系和 F^-（不含 λ）的品系接合时，溶源化状的噬菌体从供体一旦进入受体，由于不受 CI 蛋白的抑制而迅速进入裂解周期，使受体细胞破裂，释放大量噬菌体，此现象称合子诱导。

遗传学领域诺贝尔奖名录

姓名	国籍	获奖时间	功　绩
Morgan, T. H.	美	1933	发现连锁定律，奠定了遗传的染色体学说
Muller, H. J.	美	1946	用 X 射线诱导突变
Beadlr, G. W.	美	1958	建立"基因-酶"学说
Tatum, E. L.	美	1958	建立"基因-酶"学说
Ederberg, J. L.	美	1958	建立细菌杂交的方法并发现转导
Ochoa, S.	美	1959	发现多聚核苷磷酸化酶
Kornberg, A.	美	1959	发现 DNA 聚合酶
Watson, J. D.	美	1962	建立 DNA 的双螺旋模型
Crick, F. H. C.	英	1962	建立 DNA 的双螺旋模型
Wilkins, M.	英	1962	为双螺旋模型提供了 DNA 的 X 射线衍射图
F. Jacob, M.	法	1965	提出乳糖操纵子模型
Monodb, J.	法	1965	提出乳糖操纵子模型
Lwoff, A.	法	1965	揭示了溶原现象的本质
Rous, F. P.	美	1966	发现鸟类中的致瘤病毒
Holley, R. W.	美	1968	确定苯丙氨酸 tRNA 的序列、结构和反密码子
Nirenberg, M. W.	美	1968	破译遗传密码
Khorana, H. G.	美	1968	破译遗传密码
Hershey, A.	美	1969	用噬菌体感震实验证实 DNA 是遗传物质
Elbruck, M. D.	美	1969	证实噬菌体的遗传重组
Luria, S.	美	1969	论证噬菌体发生的突变
Ulbecco, R. D.	美	1975	发现肿瘤病毒
Temin, H.	美	1975	发现 RNA 病毒的反转录酶
Baltimore, D.	美	1975	发现 RNA 病毒的反转录酶
Arber, W.	瑞士	1978	提出限制与修饰，并发现Ⅰ类限制性酶
Smith, H.	美	1978	转化Ⅰ类限制酶并阐明其性质
Nathans, D.	美	1978	制作酶切图谱
Gilbert, W.	美	1978	建立化学法测序
Sanger, F.	英	1980	建立"加""减"法测序
Berg, P.	美	1980	建立体外重组技术
Benacerraf, B.	美	1980	发现细胞表面调节免疫反应的遗传决定结构
Dausset, J.	法	1980	发现细胞表面调节免疫反应的遗传决定结构
Snell, G.	美	1980	发现细胞表面调节免疫反应的遗传决定结构
Klug, A.	英	1982	对核小体结构的研究
Mcclintock, B.	美	1983	发现转座和转座因子
Kohler, G.	英	1984	建立单克隆抗体制备技术
Milstein, G.	英	1984	建立单克隆抗体制备技术
Tonegawa, S.	日	1987	证明了免疫球蛋白的体细胞重组
Varmus, H. E.	美	1989	发现鸟类肉瘤病毒中有致癌基因
Bishop, J. M.	美	1989	发现鸟类肉瘤病毒中有致癌基因
Cech, T. R.	美	1989	发现 RNA 的自体拼接

续表

姓名	国籍	获奖时间	功　绩
Atman, S.	美	1989	发现 RNA 的自体拼接
Mullis, K.	美	1993	建立 PCR 技术
Smith, M.	美	1993	建立定点突变技术
Sharp, P. A.	加	1993	发现断裂基因
Roberts, R. J.	美	1993	发现断裂基因
Gilman, A.	美	1994	发现细胞信息传递中的 G 蛋白
Rodbell, M.	美	1994	发现细胞信息传递中的 G 蛋白
Christine, N-V.	德	1995	研究了果蝇的早期发育
Edward, B. L.	美	1995	研究了果蝇的早期发育
W. Eric, F.	美	1995	研究了果蝇的早期发育
Prusiner, S. B.	美	1997	朊蛋白构象的研究
Boyer, P. D.	美	1997	三磷酸腺苷的形成过程
Wolker, J. E.	英	1997	三磷酸腺苷的形成过程
Skou, J. C.	丹	1997	发现维持细胞中钠离子和钾离子浓度平衡的酶
Blober, G.	美	1999	信号假说的建立

遗传学大事年表

1665　Hooke, R. 发表了著作《显微摄影》，首次描述了细胞。

1735　Linne, C.V. 出版了《自然系统》巨著，首创物种的二名法。

1798　Malthus, T.R. 发表了《人口论》，再这一论著的影响下，1838 年达尔文提出了“物竞天择，优胜劣汰”的学说

1806　Vauguelin 和 Robiquet 首次分离出第一个氨基酸——天冬酰胺。

1809　Lamarck, J.B. 提出“物种进化”，“用进废退”和“获得性状可以遗传”。

1838　Schleidn, M.J. 和 Schwann 发表了细胞学说。

Mulder, G.J. 发现了蛋白质。

1855　Virchow, R.L.C. 提出细胞仅来自细胞的细胞系学说。

1859　Darwin, C. 出版巨著《物种起源》，提出了著名的进化论。

1866　Mendel, G. 发表了《植物杂交实验》的论文，开创了遗传学。

1868　Darwin, C. 提出了遗传的“泛生论”。

1868　Miescher F. 发现核质

1869　Galton, F. 提出“融合遗传”说，为数量遗传奠定了基础。

1883, 1885　Weismann, A. 提出了“种质论”.

1873　Schueider, A. 首先描述有丝分裂过程。

1875　Hertwig, O. 发现受精核是双亲提供的雌雄核的融合。

1879　Flemming, W. 发现有丝分裂是染色体纵裂为二，分别移向两极。

1881　Balbiani, E.G. 在摇蚊幼虫的唾腺中发现巨型染色体。

1882　Flenuming, W. 发现灯刷染色体，并为有丝分裂定名。

1883　Van Benedin, E. 发现卵子中的染色体数是体细胞的一半。

Roux, W. 推测染色体携带遗传因子。

1884　Strasburger, E. 描述种子植物的受精作用。

1887　Weismann, A. 推测在有性生殖的生物中，染色体数目周期性减半。

1891　Boveri, T.　描述减数分裂，联会过程。

Henkin, H. 发现 X 染色体。

1895　Roentgen, W.C. 发现 X 射线。

1897　Buchner, E. 发现酶。

1898　Benda, C. 发现线粒体。

1900　De Vries, H., Correns, C. 和 Tchermak 发现孟德尔的论文，即孟德尔定律的第二次发现。

1901　De Vries, H. 提出“突变”这一概念。

1902　Mc Clung, C.E 发现性染色体。

Sutton, W.S. 和 Boveri, T. 同时提出遗传的染色体学说。

1902 ~ 1909　Batson, W. 起名“遗传学”，并创用“等位基因”、“纯合体”、“杂合体”、“杂交一代”、“上位基因”等名词及概念。

1906　Batson, W. 和 Punnet, R.C. 提出“相引”和“相斥”的概念。

Doncaster, L. 和 Raynor, G.H. 发现性连锁。

1907　Harrison, R.G. 发明组织培养。

1908　Hardy, G.H. 和 Weinberg, W. 提出哈迪 - 温伯格平衡定律。

Nilsson, E.H. 对数量性状提出多基因学说

1909　Johannsenm, W. 将遗传因子定名为“基因”，并提出“基因型”和“表型”的概念。

Garrod, A.E. 发表“先天性代谢缺陷”，开拓了生化遗传。

Janssens, F.A 发现染色体交叉的现象。

1910　Morgan, T.H. 发现连锁定律。

1911　Rous, P. 发现鸟类的致瘤病毒。

1912　Morgan, T.H. 发现雄果蝇的完全连锁。

1913　Sturtevant, A.H. 创用三点测交并建立连锁距离的概念。

Bridges, C.B. 发现减数分裂中染色体不分离现象，确证遗传的染色体学说。

1915　Twort, F.W. 分离病毒。

1917　Bridges, C.B. 发现果蝇染色体的缺失。

Carothers, E.E. 发现蚱蜢异型染色体，证明染色体的自由组合。

1918　Muller, H.J. 发现平衡致死。

1919　Bridges, C.B. 发现染色体的重复。

1920　Blakeslee, A.F. 在曼陀罗中发现各种三体，证明各染色体遗传效应不同。

1921　Goldschmidt, R.B. 首次发表有关工业黑化和生物演化的关系。

1923　Bridges, C.B. 发现染色体易位。

1924　Fenlgen, R. 和 Rossenbeck, H. 发明了 DNA 特异性 Feulgen 染色法。

1925　Sturtevent, A.H. 分析果蝇的棒眼基因，发现位置效应。

Bernstein, F. 对 ABO 血型提出复等位基因控制的学说。

1926　Sumner, J.B. 首次制备酶的结晶，并证明酶是蛋白。

Sturtvant, A.H. 发现染色体倒位现象。

1927　Muller, H.J. 以 X 射线诱发突变。

1928　Griffith, F. 发现肺炎双球菌的"转化"现象。

1929　Fleming, A. 发现青霉素。

1930　Fisher, R.A. 发表《自然选择的遗传理论》，奠定了群体遗传的基础。

1931　Crighton, H.B. 和 McClintock, B. 以玉米为材料证明了染色体的交换重组。

Stern, C. 以果蝇为材料证明了染色体的交换重组。

Wright, S 发表了《 孟德尔群体中的进化 》，构成群体遗传的部分基础理论。

1933　Painter, T.S. 开始研究果蝇的唾腺染色体。

Tiselius, T.S 发明电泳的方法。

1934　Schlesinger, M. 证明噬菌体由 DNA 及蛋白质所组成。

1935　Beadle, G.W. 和 Ephrussi, B. 揭示果蝇眼色遗传的生化机制。

1936　Stern, C. 发现有丝分裂重组。

1937　Dobzhansky, T 发表《遗传学与物种起源》一书。

1938　Sonneborn, T.M. 发现草履虫的杀伤因子。

洛克菲勒基金会主席 Warren Weaver 在理事会年终报告中首先提出"分子生物学"这一概念。

Astburu, W.T. 和 Bell, F.D. 首次发表脱氧核糖核酸的 X 光衍射研究。

Bernal, J.D. , Fankuchen, I. 和 Perutz, M. 发表胰蛋白酶和血红蛋白的 X 光衍射研究。

1939　Ellis, M.L. 和 Delbruck, M. 揭示了噬菌体的生长情况。

1940　Pauling, L. 和 Delbruck, M. 发表论文阐明量子力学理论在分子生物学研究中的作用。

1941　Beadle, G.W. 和 Tatum, E.L. 发表了"一基因—酶"学说。

Bernal, J.D 等对植物病毒进行了 X 光晶体衍射的研究。

1942　Luria, S.E 等利用电镜观察了噬菌体的形态和特征。

1943　Luria, S.E. 和 Delbruck, M. 论证抗性是细菌的固有特征，而不是诱变产生的。

1944　Avery, O.T. 等进行了体外转化实验，证实遗传物质是 DNA 而不是蛋白质。

Schriodinger 发表了《什么是生命》一书，使很多物理学家进入遗传学的研究领域，促进了遗传学的迅猛发展。

1945 Luria,S.E. 论证了噬菌体的突变,使噬菌体成为研究遗传学的重组材料。

1946 Lederberg, E. 和 Tatum, E.L. 利用细菌杂交而获得了细菌的重组子。
并证实了噬菌体的遗传重组。

1948 Boivin,A.,Vendrely,R. 和 Vendrely,C 提出了"Boivin－Vandrely"法则,即二倍体含有的 DNA 量是单倍体的二倍。

1949 Pauling, L. 等提出镰刀形贫血是一种"分子病"。
Hershey, A, D. 等进行了噬菌体的遗传重组实验。

1950 Chargaff, E. 提出任何细胞中 A 和 T,G 和 C 的碱基比率是恒定的。否定了四核苷酸理论。
研究了核酸的化学性质及酶液降解机理。

1951 McClintock, B. 首先发现了玉米中的"跳跃基因"。
Pauling,L. 和 Corey,R.B. 提出蛋白质结构的 α－螺旋模型。
Sange, F. 和 Tuppy, H. 用纸层析首次分析了胰岛素的氨基酸的顺序。

1952 Hershey, A.D. 和 Chase, M. 用标记噬菌体感染实验再次证实遗传物质是 DNA,而不是蛋白质。
Zinder, N.D. 和 Lederberg, J. 发现了广泛性转导现象。
Briggs,R. 和 King,T. 将蛙囊胚细胞核转植到去核的蛙卵内,表明移植核能进行分化。
Lederberg, J 和 Lederberg,E.M 建立了细菌的影印培养和间接选择的方法。

1953 Watson, J.D. 和 Crick, F.H.C. 建立了 DNA 的双螺旋模型。
Franklin, R 和 Wilkins, M.H.F. 拍摄到清晰的 DNA 晶体的 X 衍射照片。

1954 Gamow,G. 首先发表破译遗传密码的研究论文,提出脱氧核糖核酸和蛋白质结构具有潜在的相关性。推测遗传密码是三联密码。
Zamecnik, P.C. 和 Keller, E.B. 建立了蛋白质体外合成技术。

1955 Green, D.W., Ingram, V. 和 Perutz, M.F. 用 X 射线衍射晶体分析研究了血红蛋白的结构。
Benzer, S. 建立了染色体的精细作图方法和互补测验。提出了"顺反子"概念。
Ochoa, S. 等首次发现了多核苷酸磷酸化酶并体外合成多聚核苷酸。
Crick,F.H.C. 提出变性模板与转接器假说,预见了 tRNA 的存在。

1956 Volkin, E. 和 Astrachan, L. 首次发现噬菌体 T2 的 mRNA。
Gierer, A. 和 Schramm, G. 发现烟草花叶病毒的遗传物质是 RNA。
Tjio, J.H. 和 Levan A. 证实人的染色体数目为 2n = 46,而不是 48。
Puck,T.T. 等首次成功地培养了人类细胞。
Sarker, P. 和 Stebbins, G.L. 确定了小麦 B 染色体组的来源。

1957 Brenner,S. 根据已知氨基酸的顺序确定三联密码是不重叠的。
Kendrew, J.C. 等用 X 光衍射建立了肌红蛋白分子的三维结构模型。
Ingram, V. 确定了镰刀型贫血是由于单个碱基突变所致。
Franenkel-Conrat, H. 及 Williams, R.C. 构建烟草镶嵌病毒,证实其遗传物质是 RNA,而不是蛋白。

1958 Crick 提出"转接器"分子含有核苷酸,可以和 mRNA 精确配对。
Hoagland, M.B., Stephenson, M.L., Scott, J.F., Zamecnik, P.C. 等首次发现 tRNA。
Meselson, M. 和 Stahl, F.W 用同位素标记高速离心证实 DNA 的半保留复制。
Kornberg,A. 小组发现了 DNA 聚合酶。
Taylar, T.H. 等用放射自显影技术证实真核生物染色体的半保留复制。

1959 Ford, C.E., Jones, K.W, Briggs, J.H. 等发现 Turner 综合征的女性为 XO 型。
Jacobs, P.A. 和 Strong, J.A. 证明 Klinefelter 综合征为 XXY 型。
Lejeune, J.,Gautier, M. 和 Turpin, R. 证实 Down 综合征为 21－三体。
Weiss, S.B. 和 Gladstone, L. 建立合成核糖核酸的真核系统。

1960　Doty，P.,Marmur，J. 和 Eigner，J. 三个小组的研究同时发现 DNA 的变性与复性。
Kendrew,J.C. 等用 X 光衍射研究了血红蛋白的结构。
Perutz,M.F. 等用 X 光衍射研究了肌红蛋白的结构。
Weiss，S. 和 Hurwitz 两组同时发现 RNA 聚合酶。

1961　Jacob，F. 和 Monod，J. 提出了操纵子模型。
Brewner，S.,Jacob，F. 和 Meselson，M. 发现细菌的 mRNA。
Hall，B.D. 和 Spiegelman，S. 证实了 T_2 – DNA 和 T_2 – mRNA 的互补性。
Crick，F.H.C.,Barrett，L.,Brenner，S. 等证实了遗传密码为三联密码。
Nirenberg，M.W. 和 Matthei，J.H. 建立了无细胞系统，初步破译了 AAA、UUU、GGG 和 CCC 四个密码子。
Von Ehrenstein，G. 和 Lipmann，F. 证明遗传密码的普遍性。
Lyon，M.F. 及 Russell，L.B. 分别提出雌性哺乳动物中有一条 X 染色体失活的假说。
Beermann,W. 证明摇蚊唾腺染色体上的疏松区是基因所在的座位。

1962　Arber，W.,Dussoix，D. 发现了 DNA 的限制与修饰。
Benzer 实验室的 Chapeville，F.，Lipmann，F，Von Ehrenstein，G. 发现 tRNA 所携带的氨基酸的特异性和 tRNA 与 mRNA 结合的特异性无关。
Gierer，A.；Warrur，J.R. 及 Stachelin 三个课题组分别发现多核糖体。

1963　Monod，J.,Changeus，J – P 和 Jaconb，F 提出蛋白变构理论。
Merrifield，R.B. 合成固相肽。

1964　Yanofsky,C. 等通过对 *E.coli* 的 Trp 合成酶的研究建立了基因和蛋白质的共线形关系。
Nirenberg 等通过三联体结合实验破译全部有义密码子。
Holliday,R.　提出了 DNA 重组模型。

1965　Holley，R.W. 等发表酵母丙氨酸 tRNA 的完整序列。
Weigert，M.G. 和 Garen，A. 发现终止密码子 UAG 和 UAA。
Brenner,S.、Stretton A.O. 和 Kaplan，S. 证明 *E.coli* 中无义密码子的碱基组成。
Khorana,H.G. 通过合成重复共聚物破译全部有义密码子。
Ritossa,F.M. 和 Spiegelman 证明果蝇的 rRNA 基因在 X 和 Y 染色体的核仁形成区。

1966　Gest，J.R.，和 Yanofsky，C. 建立 *E.coli* 遗传图谱的遗传方向。
Crick，F.H.C. 提出反密码子的摆动学说。
Gilbert，W. 和 Buller-Hill，B. 分离纯化了 lac 阻遏物。
Adams,J.M. 和 Cappechi，M..R. 发现各种蛋白质的合成的起始氨基酸都是甲酰 – 甲硫氨酸。
Bretscher，M.S. 和 Marcker，K.A. 发现现核糖体上的 P 位点和 A 位点。
Weiss，B. 和 Richardson，C.C. 分离出连接酶。
Sinclair，J.H. 和 Stevens，B.J. 发现线粒体具有环状 DNA。

1967　Phillips，D.C 等对溶菌酶进行晶体学研究，分辨率达到 2Å。
Jacobson，C.B. 及 Barter，R.H. 用羊膜穿刺术开拓了产前诊断。

1968　Donahue，R.P. 等首次用细胞学方法将“达菲”(Duffy)血型基因定位于第一号染色体上。

1969　Gall，J.G. 和 Pardue,发展了原位杂交技术。
Boone，C. 和 Ruddle，F. 用仙台病毒使人、鼠细胞融合，

1970　Temin，H.M.,Mizutani，S. 和 Baltimore，D. 同时发现了 RNA 反转录酶
Smith，H.O. 和 Welcox，K.W. 发现了限制酶Ⅰ。
Gilman，A. 和 Rodbell，M. 发现传递信号的 G 蛋白及其功能。
Khorana，H.G. 等合成了酵母丙氨酸 tRNA 的基因。

1971　Danna，K. 和 Nathans，D. 绘制了第一张限制性酶切图谱。
O'Riordan,M.L. 等用荧光带技术鉴别出费城染色体是 22 号染色体缺失了 1/3 而形成。

1972 Jackson，D.A.，Symons，R.H. 和 Berg，P. 创建 DNA 体外重组。

1973 Cohor，S.R.，Chang，A.C.Y.，Boyer，H.W. 和 Holling，R.B. 首次在体外构建具有功能的细菌质粒。

1974 Berg，P 等一批著名的科学家提出暂缓进行体外重组的倡议书。

1975 Southern，E.M. 建立 DNA 印迹技术。

Grunstein，M 和 Hogness，D.S. 建立菌落膜上杂交技术。

O'Farrell，P.H. 建立高分辨双向蛋白电泳。

Kohler，G. 和 Milstein，C. 建立单克隆抗体制备技术。

继 Berg，P 倡议书后召开了 Asilomar 会议。

1976 Hozumi，N. 和 Tonegawa，S.，证实了免疫球蛋白形成的体细胞重组学说。

Stehelin，D.，Varmus，H.E.，Bishop，J.M. 和 Vogt，P.K 发现鸟类肉瘤病毒中含有癌基因。

Kan，W.Y.，Golbus，M.S. 和 Dozy，A.M. 首次用分子杂交技术用于遗传病(地中海贫血)的产前诊断。

1977 Berget，S.M.，Moore，C. 和 Shap，P.A 发现腺病毒头部蛋白基因中存在内含子和外显子，提出了断裂基因的新概念。

Maram，A.M. 和 Gilbert，W. 建立了 DNA 的化学测序法。

Sanger，F. 建立了 DNA 测序的"加""减"法，即酶法测序。

1978 Chang，A.C.Y.，Nunberg，T.H.，Kanfaman，R.F. 首次将真核基因 (dhfr) 在细菌中进行表达。

Collins，J. 和 Hohn，B. 建立了装配型载体，克隆了 DNA 大片段。

Hutchison，C.A.，Phillips，S.，Edgell，M.H. 建立定点突变技术。

Blackburn，E.H. 等发现了四膜虫端粒的串联重复顺序。

1979 Solomon，E. 和 Bodmen，W.P. 限制性片段长度多态性的研究。

Wang，A.H-J 和 Rich，A. 提出 Z-DNA 模型。

1980 Botstein，D.，White，R.L.，Skolnick，M. 和 Davis，R.W. 用限制性片段长度的多态性构建人类遗传学连锁图。

1981 Cech，T.R.，Zaug，A.J. 和 Grabowski，P.J. 发现四膜虫 rRNA 的自体拼接。

Banerji，S.，Rusconi，S. 和 Shaffner，W. 发现增强子。

Brinster 和 Palmiter 首次将疱疹病毒的 TK 基因转移到小鼠中表达。

1982 Tabin，C.J.，Bradley，S.M.，Bargmann，C.I.，等证实癌基因中单个碱基的突变而引起肿瘤的产生。

1983 Sulston，J.E.，Shierenberg，E.，White，J.G 和 Thomson，J.N. 建立了线虫(*C.elegans*)胚细胞谱系。

Herrera-Estrells，L. 等应用 Ti 质粒将基因转入植物。

1984 McGinnis，W. 等发现果蝇的同源异型盒及同源异型基因。

Schwartz，D.C. 和 Cantor，C.R. 建立了脉冲变角凝胶电泳技术，用于大片段 DNA 的分离。

1985 Mullis，K. 等建立了 PCR 技术进行体外扩增。

1986 Benne，R. 等发现 RNA 编辑的现象。

Symons，R.H. 对植物类病毒拟病毒提出了锤头结构模型。

1987 Burke，D.T.，Carle，G.F. 和 Olson，M.V.，构建了酵母的人工染色体 YAC。

Hoffman，E.P.，Brown，R. 和 Kuakel，L.M. 鉴定了杜兴氏肌肉萎缩症突变基因的产物。

Hoopoer，M.，Hardy，K 等用胚的干细胞(ET)进行制备转基因鼠取得成功。

Kuehb，M.R.，Bradly，A. 等将 HPRT 突变基因导入小鼠，建立了尼—莱氏综合征的动物模型。

Greider，C. 和 Blackburn，E.H. 分离了端粒酶。

1988 Whyte，P.，Buckkovich，K.J.，Horowitz，J.M 等研究了 Rb 抗癌基因的功能。

1989　Rould，M.A. 等报导了 *E.coli* 中氨基酰 .tRNA 合成酶的结构和功能。

1990　Blackburn，E.H. 等建立了端粒的复制模型。

Sinclair，A.H 等克隆并证实人类和哺乳动物性别决定的主要基因 *SRY/sry*。

1991　Blum，B.，Sturm，N.R.，Simpson，A.M. 和 Simpson，L. 提出了 RNA 编辑的转酯反应模型。

1992　Momand，J，Seto，E. 和 Donehower，L.A. 三个实验室揭示了 *p53* 的抗癌作用机制。

1993　Luan，D.D 等提出了长分散序列的反转录机制也是归巢的一种机制。

1994　Perler，T.B.，Davis，E.O. 等提出"内蛋白子"的概念。

Debyser，Z. 等提出细菌复制叉上的复制体模型。

1995　Zimmerly,S. 等发现Ⅱ类内含子的结构及剪接。

1996　Dietrich，W.F 等绘制了小鼠基因组的完整遗传图谱。

Dib,C 等在 5 264 个微卫星的基础上绘制了人类的完整遗传图谱。

1997　Wilmut,I 利用绵羊的乳腺细胞的细胞核成功地克隆了一只小羊。

1999　国际人类基因组计划联合研究小组完整地破译出人类第 22 号染色体的遗传密码。

2000　完成了人类第 21 号染色体的测序。5 月宣布国际人类基因组计划完成时间将再度提前。预计从原定的 2003 年 6 月提前到 2001 年 6 月。2000 年 6 月 26 日，美，英，日，德，法，中六国政府和有关科学家共同宣布人类基因组工作草图绘制成功。

果蝇和玉米的遗传图

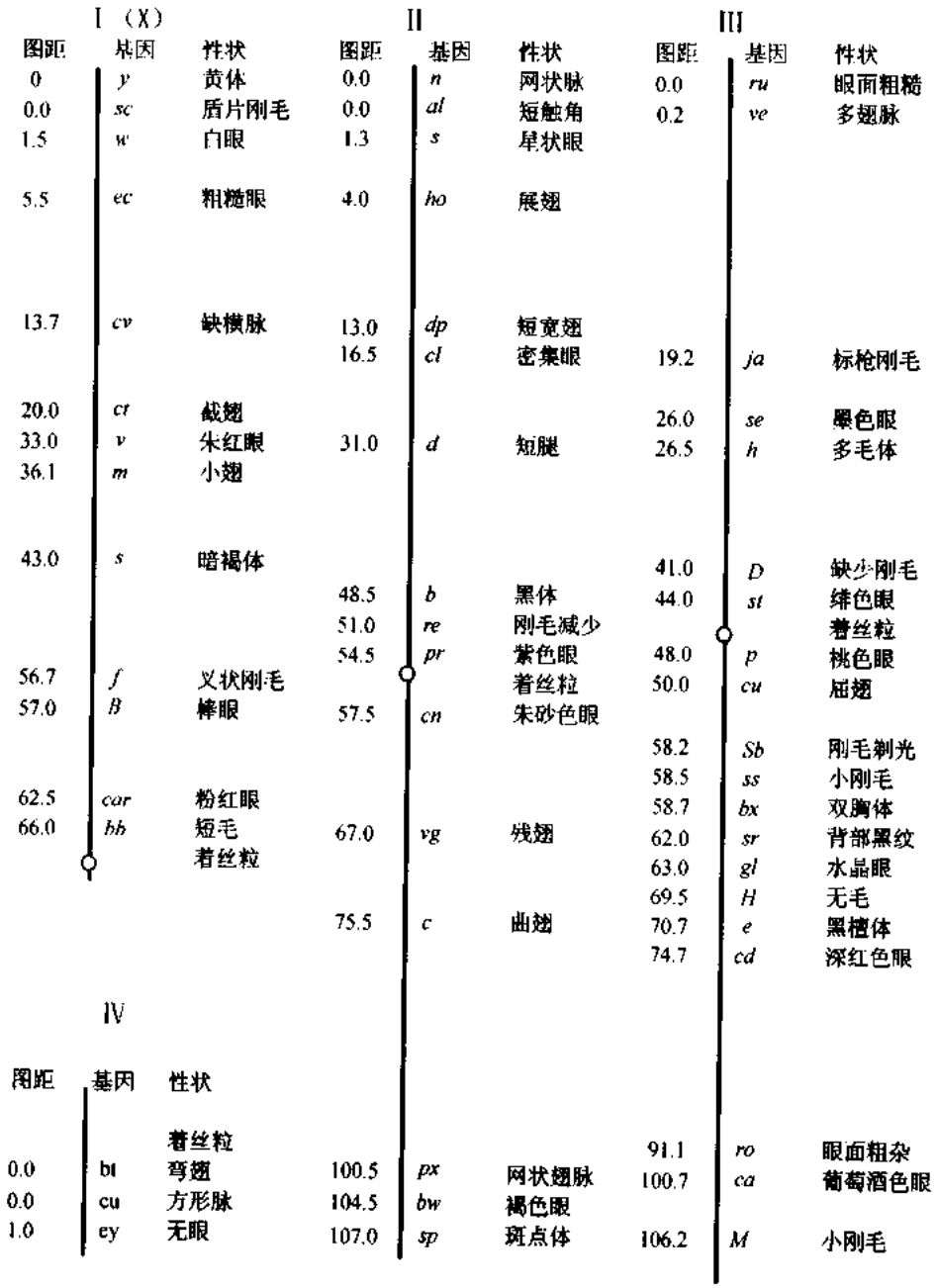


图 1 果蝇（*Drosophila melanogaster*）的遗传图

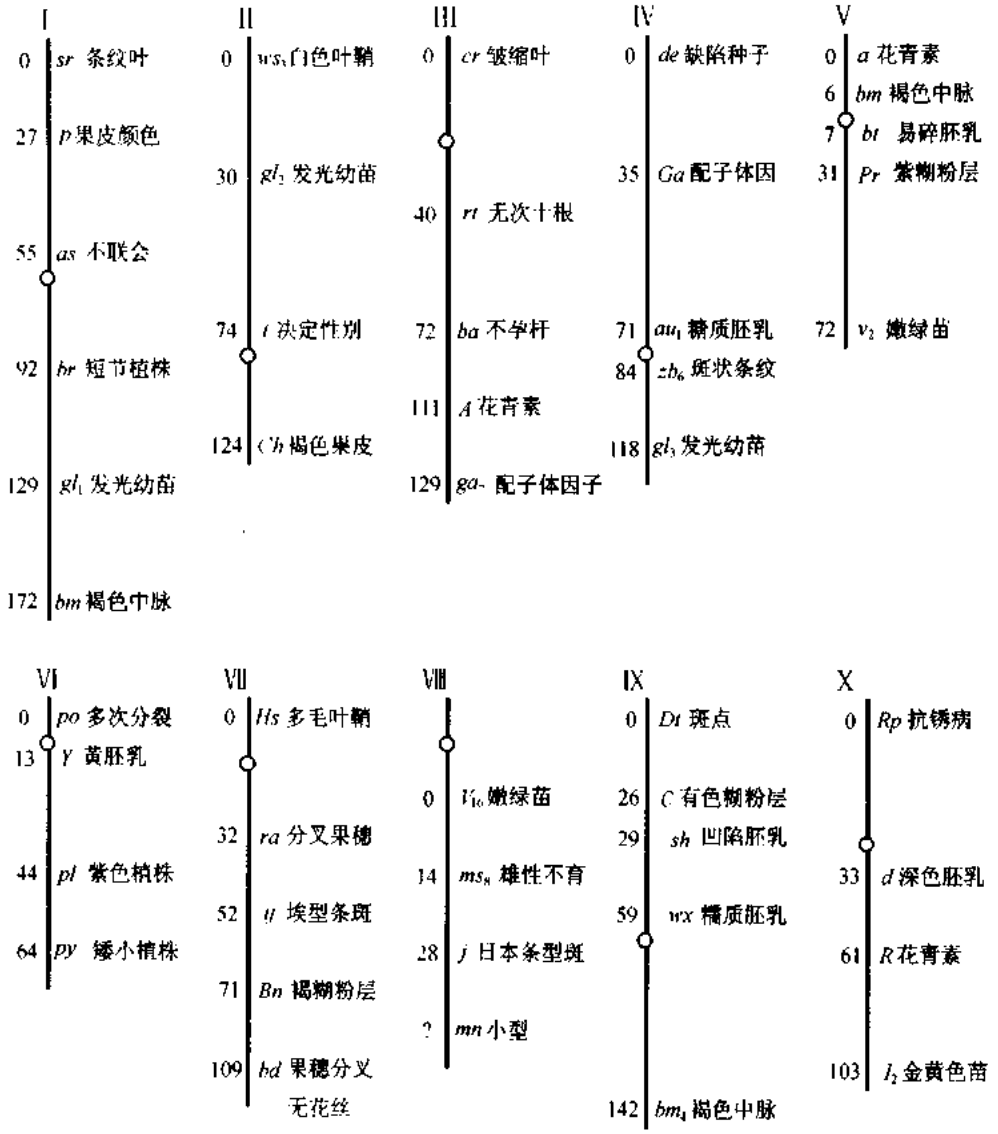

I
0 sr 条纹叶
27 p 果皮颜色
55 as 不联会
92 br 短节植株
129 gl1 发光幼苗
172 bm 褐色中脉
II
0 ws1 白色叶鞘
30 gl2 发光幼苗
74 t 决定性别
124 C1 褐色果皮
III
0 cr 皱缩叶
40 rt 无次生根
72 ba 不孕杆
111 A 花青素
129 ga- 配子体因子
IV
0 de 缺陷种子
35 Ga 配子体因
71 au1 糖质胚乳
84 zb6 斑状条纹
118 gl3 发光幼苗
V
0 a 花青素
6 bm 褐色中脉
7 bt 易碎胚乳
31 Pr 紫糊粉层
72 v2 嫩绿苗
VI
0 po 多次分裂
13 Y 黄胚乳
44 pl 紫色秸株
64 py 矮小植株
VII
0 Hs 多毛叶鞘
32 ra 分叉果穗
52 tl 埃型条斑
71 Bn 褐糊粉层
109 bd 果穗分叉
无花丝
VIII
0 V16 嫩绿苗
14 ms8 雄性不育
28 j 日本条型斑
? mn 小型
IX
0 Dt 斑点
26 C 有色糊粉层
29 sh 凹陷胚乳
59 wx 糯质胚乳
142 bm4 褐色中脉
X
0 Rp 抗锈病
33 d 深色胚乳
61 R 花青素
103 I2 金黄色苗

图 2　玉米的遗传图

索 引

-10 顺序 299
13-三体(Patau 综合征) 691
14-三体 677,691
18-三体(Edward 综合征) 691
1 型帽子 (cap 1) 334
Ⅰ类酶(Gln-tRNA 合成酶) 392
Ⅰ类内含子 341
Ⅰ型锌指 575
2′-3′环磷酸 327
21-单体 677,679
21-三体 679,691
2 型帽子(cap2) 334
2 型内部启动子 317
Ⅱ类酶(Asp-tRNA 合成酶) 392
Ⅱ类内含子 344
Ⅱ型锌指 576
-35 顺序 299
5′-OH 527
5P- 综合征 668
5-氮胞苷 572
5-溴脱氧尿嘧啶(5-Bromodeoxy uridine, BrdU) 236
6-磷酸葡糖脱氢酶(G-6-PD) 564
A(atachment)蛋白基因 494
Ac-Ds 系统 735
AC-S(achaete-scute 无刚毛基因的产物) 577
ADPR(adenosine diphophate tibosyl; 腺嘌呤核苷二磷酸核糖体糖基) 417
ADP-核糖转移酶(ADPRT) 481
AHC(adrenal hypoplasia) 814
Alu-等同家族 760
Alu 基因家族(Alu family) 222,760
Alu 家族(Alu family) 222,760
Alu 结构域(Alu domain) 429
ALV(avian leukosis virus ,禽白血病毒) 610
Ames 测验(Ames test) 651
AMH(抗副中肾激基基因,也称副中肾抑制基因 mullererian inhibiting substance, MTS) 814
ANT-C(触角足复合体) 796
ANT-C 复合座位 796
Antp(触角足 antennapedia) 797
AP-1 608
AP 位点 658
ARS 元件 288
att B 161,714
att P 161,714
attλ 161,714
A 盒(boxA) 510
a 特异基因(asg) 545
A 位点(进入位点,entrysite) 386,397
a 型 537
A-型周期蛋白 588
a 因子 537
B1 家族 223
Barry Holland 435
bcr(断裂点簇区域,break-point claster region) 613
Bicoid 基因 780
BiP 436
BOB′ 714
boxA 316
boxB 317
boxC 316
Brenner 362
B1 家族 760
B 盒(boxB) 510
B 淋巴细胞 550
B-染色体 32
B-型周期蛋白 588
C. elegans 586,595,771
CI(clear plaque)基因 512
C Ⅰ 505
C Ⅱ 蛋白 505,527
C Ⅲ 505
C Ⅲ 蛋白 505,526
c-*abl* 613
CAD 基因 537
CAF-1 283
CAK(cdc2-activation kinase) 598
CAP(catabolite gene activator protein) 463
CBP(cap binding protein) 407
CD (Campomelic dysplasis)症 814
CEN 顺序 204
c-*erb* 603
c-*fos* 609
C-Fos 577
Chi 709
chi(十字型) 709,718
Chi 序列 709
C Ⅰ (Clear platne) 基因 507,512
cin 基因 718
c-Jun 577
CIB 技术 674
Col 质粒 193

copia 反转录子　756
cp(coat protein)外壳蛋白基因　494
cro　505
Cro 蛋白　524
CsCl 密度梯度离心　621
CSF(cytosstatic factor 细胞静止因子)　590
c-sis　604
C-带　212
C 值 (C Value)　226
din(damage inducible)　663,718
Dmc1　709
Dna A　258
Dna B　258
Dna C　258
Dna K/DnaJ　258
DNA pol Ⅱ　246
DNA pol Ⅲ　246
DNA pol Ⅲ全酶　251
DNA polα　279
DNA polδ　279
DNase I 的超敏感位点　543
DNase Ⅰ　543,567
DNA 长度的多态性(DNA length polymorphisms)　850
DNA 聚合酶Ⅰ(DNA polymerase Ⅰ,DNA pol Ⅰ)　246,247
DNA 膨松(DNA puffs)　532
DNA 糖基酶(glycosylases)　658
DNA 指纹(DNA fingerprints)　224
DSS(剂量敏感性反转基因 dosegesencitive sex reversal)　815
dut-突变体(dUTPase 缺失)　274
D 环　242
D 环(D arm)双氢尿嘧啶　376
D 基因 552
E12/E47　577
EF－G　412
EF-Ts　412
EF-Tu　412
EGF 受体 (EGFR)　603
eIF-2　406
eIF-3　406
eIF-4A　406
eIF-4 因子　407
eIF-6　406
E_L(near HML)　543
Ellobius 属的鼹－田鼠(mole-vole)　813
En 因子　739
env　745
EnvZ　495
E_R(near HMR)　543
ERE 序列(雌激素的靶序列)　576
ESP(eIF-2-stimulating protein)　585
E 沉默子　543
E 位点　386,397
Fabry 病　653
FB(折回,foldback)因子　757
FB 家族　757
F 纤毛(F pill)　141
F 因子 (F-factor)　141,265
F 因子复制起点 *oriT*　141
F 因子复制起点 *oriV*　141
G(－)　485
G(＋)　485
G1 周期蛋白　588
gag　745
GAL4　312, 578
GAL80　578
GAP(GTPase actiration protcin,GTP 酶活化蛋白)　606
Garrod. A　115
GCN4　312
gin(G segment inversion)　485,673,718
Goodfelow　812
GO 系统　659
GRE(糖皮质激素反应元件,glucocorticoid response element)　574
GRE 序列(糖皮质激素的靶序列)　574
GroEL　433
GT-AG 法则(GT-AG rule)　338
GU-AG 法则　346
G 带(G banding)　211
G 蛋白　539
G 片段　485,674,718
H0　545,548,697
H_2A·H2B 二聚体　278
H3 mRNA　320
$H3_2$·$H4_2$ 四聚体　281,283
Haldane 的图谱函数(Haldane's mapping function)　80
HCR(控制血红素阻遏物 hemim-controlled repressor,又称 EIF-2 激酶)　585
HHG(hypogonadotrphic hupogonadism)　815
Hin　718
hin(h segment inversion)　485
himA　663,714
himD　714
hin 基因　673,718
HLH(helix-loop-helix)　577,807
HlyA　435
HlyA 的分泌模型　435
HlyB　435
HlyC　435

HlyD 435
HMG box 812
HMG 539,568
HMG 超家族 814
HML 539
Holliday 模型 697
Hs 基因 545,718
HSE(热休克反应元件,heat shock response element) 574
HU 256
Hu 蛋白 190
I(inducer) 型雄果蝇 740
Ig 基因的 V.C 重组 612
Ig 轻链 553
Ig 重链 553
IHF 714
int(integrase) 506,523,527,714
I Q 113
I 沉默子 543
I 类内含子 339
Kenya 珠蛋白 671
Klenow 片段 247
K 轻链 549
L.Barr 560
lacA 446
lacI 446
lacY 446
lacZ 446
Lac 阻遏物(lac repressor) 447
LCI(light chain) 582
LDH-C 852
Lepore 珠蛋白 671
LexA 659
Li-Fraumeni 综合征 599
lin 14 586
lin 4 586
lys(lysis) 蛋白基因 493
L 型结构 378
malE 基因 502
malF 基因 502
Mary Mitchell 621
MAT(matimg) 座位 537
McClintock 714
McClintock 730
MHR 537
mic F 495
MRE(金属反应元件,metal response element) 574
MSH4 (mismatch-repair) 709
MuA 转座酶 728
Muller 206,674
MutH 661
MutL 661
MutS 661
Mu 噬菌体 239,718
myc 基因 586
M 期促发因子(M-phase puomoting factor,MPF) 291,580,587
M 期激酶(p34-cyclin B) 588
M 期进入限制点(commitment to M phase) 587
M 细胞型 (M cytotype) 743
M 型(母本贡献的,matenal contributing) 740
M-周期蛋白 588
N 504
NAD(nicotinamide adenine dinuclestide) (辅酶 I) 417
nanos 基因 780
neu 基因 603
NPD 非亲二型(non-parental ditype) 86
ν 小体(? 钮-body 即 nu body) 197
nusA 510
nusB 510
nusE 510
nusG 510
nut(N utilization)位点 509
nutL 509
nutR 509
N 核苷酸(N nucleotide) 557
N-甲酰甲硫氨酸-tRNA(N-formyl-methionyl-tRNA) 401
N-乙酰-糖胺(N-acetyl-glucosamine) 437
OCT 元件 318
OD 值(optical density) 183
OmpC 495
OmpF 495
OmpR 495
Oriλ 258
O 蛋白 258
O-糖苷键(o-linked glycosylation) 437
O-体(O-some) 258
O 型帽子(capO) 334
P bαx 546
p21 597
p28V-sis 604
p53 蛋白 596
p53 基因 655
PABP(poly(A)结合蛋白) 585
PD 亲二型(parental ditype) 86
phr 基因 656
pol 745
pol Ⅲ′复合体 251
pol III* 252
poly(A) 尾巴 332
POP′ 714
ppGpp 499

pppGpp 499
(p)ppGpp 的合成酶 500
pQ 507
psa(photosynesizer Ⅰ) 基因 352
P 蛋白 250
P 核苷酸(P nuclectices) 557
P 盒转录因子 546
P 位点(或称供位,donor site) 386,397
P 细胞型(P cytotype) 743
P 型(父本贡献的,paternal contributing) 740
P 因子(P factors) 741
Q 505
qut(Q-utilization) 507
Q-带 212
R(reactive) 型雌果蝇 740
Rad51 509
RAG(recombination active gene) 558
Ras-GAP 606
Ras 蛋白 605
ras 原癌基因 605
Rb 592,610
Rb(视网膜母细胞瘤,retinoblastoma) 596,610
RecA 662,710
recA 基因 662
RecBC 710
recBC 基因 662
RecD 710
recF 662
RecG 713
rec 途径 662
RelA 500
relA 基因 500
relC 500
Rep(replicase) 基因编码 493
RF-A 285
RF-C 285
Rhyncosciara 532
Rh 血型 66
RNA Ⅰ 292
RNA Ⅱ 292
RNase Ⅲ 325,527
RNase D 322
RNase H 292,748
RNase Ⅱ 501
RNA 的转换率 584
Rop(或 Rom) 292
rpIK 497
rplJ(L10) 330
rplL(L7/L12)(核糖体大亚基蛋白) 330
rpoB(RN 聚合酶? 亚基) 330
rpoC 330
RuvAB 713
ruvC, 713
R-带(R-Bands) 214,218
r-蛋白(r-protein) 383
r-蛋白的自我调控 490
R-环定位法(R-loop mapping) 187
R 因子(resistance factor) 146
s RNA 功能域(s RNA domain) 429
S. Wright 832
S. Wright 效应 832
S1 酶 258
SC 组型(Karyotype) 29
S-D 顺序(Shine-Dalgarno sequence) 399,409
SecA 433
SecB 433
SecE/Y 433
SF-1(steroidogenic factor-1) 813
SH2(肉瘤基因同源区 2,src homology region-2) 602,613
SIR(silent information regulator) 312,543
sm 结合位点 348
snRNP(U1,U2,U4,U5,U6) 346
SOS 反应(SOS response) 663
SOS 修复系统 663
Southern 杂交 185,567
Sox9(SRY-相关基因,SRY-related box) 814
spm- 依赖性座位(spm- dependent locus) 735
Spm 因子 739
spm-可抑制基因座(Spm- suppressible locus) 740
spoT 500
src 基因 602
SRE(血清反应元件 serumresponse element) 574
Sry 812
SRY(sex-determining regicn of Y chromosoine) 45,812
SSB 254,661,712
STE 2 基因(sterile matation) 545
S 区(S regions) 561
S-腺苷基甲硫氨酸 334
S 型位置效应 672
T. Boveri 36
T2 噬菌体 155
T7 噬菌体 265,267,482
TAFs(TBP-associated factors) 312
TATA 框 309,310
TBP(TATA box-binding protin) 312,601
TCR(T cell receptor) 556
TCR 座位(T 细胞受体,T cell recepter) 612
ter 顺序 271
TFⅢ B 317
TFⅢ A 317

TFⅢC 317
TGF-β(转生长因子,transforming growth factor) 814
Ti质粒 193
Tn10转座酶 731
TnA74家族 723
tnpA基因 740
tnpB基因 740
tnpR基因 740
TopI 716
*torso*基因 781
TPK(酪氨酸蛋白激酶) 603
TRE (佛波酯效应元件(TPA response element) 574,609
TRE(肿瘤诱导剂反应元件,tumorgenic agent response element) 574
tRNA的半分子(tRNA half-molecules) 327
tRNA核苷转移酶(tRNA nucleotidyl transferase) 322
trpR调节子 461
Tu,Ts及S1蛋白 494
*TyA*区域 753
*TyB*区域 753
Tyr的自身磷酸化 602
TψC臂 376
T蛋白 286
T细胞 550
U2AF(U2 auxiliary factor) 350
U7snRNA 320
UAS_G(upstream activator sequence-galactose) 579
UBF1(上游结合因子1) 315
ung-(尿嘧啶N-糖苷酶缺失) 突变体 274
UreD 656
UvrABC内切酶 656
UvrA 656
UvrABC系统 656,658
UvrB 656
UV-复活(UV-reactivation) 663
UV光化二聚体 658
U形管实验 139
v-jun, 606
VSP(very short patch repair)系统 660
V型位置效应 672,677
W(Weigle) -复活 663
Walter Sutton 36
Warner法 111
Western杂交(Western blotting) 187
WT1基因 813
X:A 805
Xis(excisioase) 714,716
*xis*基因 714
XO型(XO type) 46
XY型(XY type) 45
X和Y连锁 53
X连锁(X-linkage) 52,53
X连锁脊髓和延髓肌萎缩(Kennedy's症) 654
X连锁显性遗传(X-linked dominant inhertance) 50
X连锁隐性遗传(X-linked recessive inhertance) 50
X染色体 39
Ya(γ-associated sequence,Y-装配顺序) 807
*Ya*和*Y* 543
Y-Z的交界处 547
Y连锁(Y-limkage) 52,53
Z-DNA模型 176
ZO型(ZO type) 46
ZW型(ZW type) 46
α/a二倍体细胞 537
α-淀粉酶 505
α-鹅膏蕈(α-amanitin) 308
α-溶血素(α-haemolysin,HlyA) 435
α-特异性基因(αsg) 541
α型 537
α因子 537
β"夹子" 251
β-半乳糖苷酶 446
β-半乳糖苷乙酰转移酶 446
β-半乳糖苷透性酶 446
β-折叠(β-Sheet) 392
γ复合体 251
θ型复制 238
λdg(λd gal) 163
λDNA 714
λ噬菌体 164,240,258
ρ依赖性终止子(Rho-dependent terminator) 304
ρ因子(rho factor) 304
σ复制 262
σ亚基 250
σ亚基的替换 475
σ因子 301
τA 318
τB 318
τ二聚体 251
φ29 DNAs 268
φχ174 260
ψ(假尿嘧啶) 376

吖啶橙(acridine orange) 649
阿德伯格(E. H. Adelberg) 145
阿拉伯糖操纵子(ara operon) 468
阿斯特布里(W. T. Astbury) 171
埃弗里(O. T. Avery) 164
埃弗斯(B. Ephrssi) 115
埃牟尔森(Emerson) 100

埃瓦姆斯(H. Evarmus)　594
癌基因(oncogene)　594,677
癌基因假说　594
艾滋病　647
艾滋病毒　647
安慰诱导物(gratuitous inducer)　449
氨基嘌呤(2-aminopurine 2-AP)　647
氨基酸臂　376
氨基酸的臂或突起(arm or protrusiton)　456
氨基酸饥饿　499
氨基酰-tRNA　390
氨基己糖苷酶 A(hexosaminidase A,Hex A)　缺乏症　125
氨甲喋呤(Methotrexate)　423,534
氨肽酶(aminopeptiase)　405,436
暗盒(black box)　学说 624
暗箱模型(cassette model)　539
暗修复(dark repair)　656
奥利维拉(Olivera)　273
奥林斯(Olins)夫妇　197
奥特曼(S. Altman) 336

八聚体(Octamer)　310
巴尔希莫(D. Baltimore) 293
巴尔小体 (Barr body) 33,195,564
靶重复　721
白喉毒素(diphtheria toxin)　417
白化病(albinism)　122
白基特(Buvkitt)　淋巴瘤 612
白剪秋罗(Lychnis senno)52
白色体(leukoplasts)　623
白眼基因(whiitee gere)　569
摆动假说(wobble hypothesis) 373
斑点杂交(dot blotting)　185
斑色染色体(Harlequin chromosome)　237
半保留模型(semioconservetive model)　234
半不连续(semidiscontinuous)　273
半合子(hemizygote)　50
半甲基化　571
半染色体转变(half-chromatid conversion)　705
半乳糖-1-磷酸尿苷转移酶(galactose-1-phosthate uridyltransr-erase Gal-1-PUT)　121
半乳糖表面异构酶(glactose epimerase ,E.)　466
半乳糖操纵子(*Gal* operon)　466
半乳糖激酶(galactokinase)　466
半乳糖血症(Galactosemia)　11,10,121
半乳糖转移酶(galactose transferase,T)　466
半上位(semiepistatic)　69
半衰期　502
半显性(semidominance)　62
伴性遗传(sex-linked inheritance)　50

棒状眼基因 B(Bareye)　78
孢子(spore)　35
孢子体(sporophyte)　35
孢子形成(sporulation)　476
胞质分裂(cytokinesis)　34
胞质基因杂合细胞(cytohet)　630
胞质连接　774
胞质桥(cytoplasmic bridges)　774
保持系(n)　rfrf　635
保守转座(conservative tranposition)　725,726
报春花属(Primula)　72
鲍林(Pauling)　171,552
鲍文奎　688
爆发式新种起源　676
贝恩 (R. Benne)　355
贝尔(Bell)　171
贝内尔奇(Benerji)　310
贝内特 (J. C. Bennett)　552
贝特森(Bateson)　10,75
背腹轴(dorsal-vontral)　776
被编辑的前体 RNA(pre-edited RNA)　359
被阻遏水平(repressed levd)表达　445
本底(basal)表达　445,449
本泽(Benzer)　130,368
苯丙酮尿症(phenylketonuria, PKU)　122
比德尔(G. Beadle) 115
必要基因(essential genes)　68
毕晓普(J. M. Bishop)　594
臂(arms)　714
臂的组件　722
臂间倒位(pericentric inversion)　673
臂内倒位(pracentric inversion)　673
边界顺序　338,346
编辑(editing)　355
编辑体模型　358
编码端(codig ends)　556
编码连接(codig joint)　556
鞭毛蛋白　484
扁囊网　426
变构调控(allosteric control)　449
变形虫(*Amoeba proteus*)　294
变性(denaturation)　180
变性定性法(denaturation mapping)　240
标记获救(marker rescue)　255
标记追踪(pulse-chase)　294
标识顺序(signature sequence)　392
标准差(standard deviation,SD)　103
标准误(standard error)　104,834
表面糖蛋白基因 VSG(variable surface glycoprotein)　351
表现度(expressivity)　58

表现型(phenotype) 10
表型模拟(phenocopy) 59
表型模拟剂(phenocopying agent) 59
并发系数(coefficidence of coincidence or coincidence) 82
病毒癌基因(virus oncogene, v-onc) 594
病毒超家族(vira superfamivly) 757
病毒样颗粒(*vLPs*, Virus-like particles) 755
波形蛋白(vimentin) 208
伯恩斯坦(Bernstein) 64
伯内特(F. M. Burnet) 155,548
不等交换(unegual crossing over) 667,851
不对称性(asymmetry) 775
不可诱导性(uninducible) 450
不连续复制(discontinuous replication) 273
不同步培养 240
不完全显性(incomplest dominance) 62,112
不完全性连锁基因 53
不稳定突变等位基因(unstable mutant allele) 734
不相容性(plasmid incompatibility) 194
不育基因(fru, fruitless) 809
不育基因(sterile gene, *STE*) 538
布莱克本(Blackburn, E. H) 288
布莱克斯利(Blakeslee, A) 37
布里格斯(Robert Briggs) 762
布鲁姆 (Blum, B) 358
布鲁斯(Burus) 145
布洛贝尔(Blobel, G) 427
部分二倍体(partial diploid) 135, 147,451
部分合子(merozygote) 半合子 149
部分退火 329

藏报春(*Primula sinensis*) 57
操纵基因(operator) 443,445,447
操纵子(operon) 443,446
草莓(*Fragaria*) 52
侧环(looping out) 570
侧体 707
侧翼成分(lateral elements) 707
侧翼序列 W. X 542
测交(test cross) 8
层连蛋白(Lamins) 590
插入序列(insertion sequence, IS) 141, 719
插入突变剂(intercalating mutagens) 649
查可夫(Chargaff) 170
差错倾向修复(Error prone repair) 663
缠绕数 T(Twisting number) 180
产前诊断(prenatal diagnosis) 125
长-补丁修复(long-patch repair) 658
长单拷贝序列(long single copy seguence, LSC) 621
长的分散序列(long interspersed sequence, LINES) 757
长的分散因子(long interspersed elements, LINEs) 223
长末端重复序列(LTR) 610,611
长期调节(long-term regulatiion) 530
长型 copia 756
常染色体(autosome) 33
常染色质 (euchromatin) 33,195,563,563
超雌(superfemale) 56
超螺旋(Supercoied) 179
超敏感区域(hypersensitive region) 567
超敏感位点(hypersensitive site) 567
超念珠(superbeads) 模型 200
超声波 125
超数染色体(supernumerary chromosome 或 B-染色体) 32
超显性(overdominance 或 Superdominance) 112,827,847
超雄(supermale) 56
超修饰(hupermodified 482
超优势度(predominance) 801
超阻碍(super-repressed) 444
沉默暗箱(silent cassettes) 539,542
沉默框(HML 和 HMR) 312
沉默突变(silent mutation) 641
沉默子(silencers) 543
成虫盘(disk) 116,766
成核作用(nucleation) 180
成环滚环复制(looped rolling circle replication) 260,264
成熟促进因子。(muturior promoting factor, MPF) 583
成熟酶(maturase) 231
成熟前起始(premature initiation) 286
赤道板(equatorial plane) 25
重叠调控 484
重叠核苷酸 487
重叠基因 (overlapping gene) 192
重复(duplication) 652,670
重复共聚物 366
重复基因(dunlicate genes) 70
重复顺序的扩增(expansion of a three-basepair repeat) 653
重建(restitution) 666
重组合(recombination) 12
重组节(recombination nodules) 707
重组率(recombination frequency) 79
重组修复(recombination-repain) 661
重组值(recombination value) 79
重组子(recombinant) 147
出口的功能区(exit domain) 386
初级精母细胞(primary spermatocytes or meiocytes) 34
初级精原细胞(primardial spermatogonia) 34
初级卵母细胞(oocytes) 34
初级卵原细胞(primary oogonia) 34
初级性分化 811
初始配对-规则基因 791

触动因子(trigger factor) 433
传递等级绘图 146
传递遗传学(transmission genetics) 816
串联体(concatemer) 272
串联重复单位 531
锤头(hammerhead) 二级结构 353
纯合子(homozygote) 10
雌性特异剪接 808
雌性途径基因(FPGs) 815
雌性异配性别 46
次黄嘌呤(H) 649
次黄嘌呤鸟嘌呤磷酸核苷转移酶(HGPRT) 124
次级精母细胞(secondary spermatocytes) 34
次级精原细胞(secondary spermatogonia) 34
次级卵母细胞(secondary oocyte) 34
次级卵原细胞(secondary oogonia) 34
次级性分化 812
次近端序列元件(proximal sequence element,PSE) 318
从头合成(de novo) 246
从性显性(sex-influenced dominance) 53
从性遗传(sex-influenced inheritance) 53
粗面内质网(rough ER,RER) 426
粗线期(pachytene) 29
促成熟因子(maturation-promoting factor, MPF) 291,580,587
促进(promote) 594
脆性X染色体综合征 653
错配修复(mismatch repair) 659
错误倾向修复(error prone repair) 646
错义突变(missense mutation) 418,640
错义抑制(missense suppressor) 381
错折叠(misflided) 436

达尔文(Charles Darwin) 840
大T蛋白 229
大孢子发生(megasporogenesis) 35
大孢子母细胞(megaspore mother cell) 35
大肠杆菌素(colicins) 193
大沟(major or wide groove) 172
大猩猩(*Gorilla gorilla*) 680
大型染色体(macrochromosome) 32
大熊猫(*Ailuro poda melanoleuca*) 218
带型(banding patterns) 209,211
戴维斯(B. D. Davis) 139
单倍化(haploidization) 93
单倍体(haploid) 29
单倍体特异性基因(hsg) 545
单极纺锤体(monopolar spindle) 46
单价体(univalent) 30,686
单孔制网 679
单链交换(single-strand exchange) 662
单链结合蛋白(single strand binding protein, SSB) 254,661
单链取代(single-strand uptake) 710
单链同化(single-strand assimilation) 710
单链线状病毒 269
单链置换(strand displacement) 268
单链转移复合体(strand transfer complex) 729
单亲遗传(uniparental inheritance) 623,630
单顺反子(monocistron) 322,771
单体(monosomy) 691
单向易位(simple translocation) 676
单向重组 698
单眼缺乏(ocelliless) 533
单一序列(Unique sequence) 219
蛋白二硫异构酶(protein disulfide isomerase,PDI) 436
蛋白分检(Protein sorting) 420
蛋白石(opal) 密码子 370,418
蛋白质的运输(trafficking) 420
蛋白质合成起始(initiation) 397
蛋白质合成延伸(elongation) 397
蛋白质合成终止(termination) 397
氮芥(NM) 649
导肽(leader sequence) 417
导肽酶(leader peptidase) 434
倒位(inversion) 592,673
倒位环 674
倒位片段(invertable segment) 485
倒位重组 718
德雷尔(W. T. Dreyer) 552
灯刷染色体 531
等臂染色体(isochromosome) 210
等上位(isoelnstatic) 70
等数分裂(equational division) 31
等位基因(allele) 10
等位基因间的互补(interallelic complementation) 452
等位排斥(allelic exclusion) 558
等位频率(allelic freguencies) 818
等位系列(allelic series) 64
等位性(allelism) 64
低频转导(low-frequency transduction, LFT) 163
低亲和性位点 458
狄夫瑞斯(Devries) 2
地钱(*Liverwort*) 621
地中海贫血(thalassemia) 128
第二次分裂分离(second division segregation) 85
第二次减数分裂 28
第二个附着位点(secondary attachment sites) 714
第二极体(second polar body) 34
第二死亡期(M2) 592
第二性比(secondly sex ratio) 50
第二性征(secondary sexual characteristics) 805

第三性比(thirdsex ratio) 50
第一次分裂分离(first division segregation) 84
第一次减数分裂 28
第一极体(first polar body) 34
第一性比(first sex ratio)50
颠换(transversion mutation) 640
点估计 21
点突变(point mutation) 640
迭氮胸苷(AZT, azidothymidine) 647
定点突变(site-specifc in vitre mutation) 650
定位因子(positioning factor) 318
动力学校对(Kinetic profreading) 394
动物园印迹(Zoo blot) 185
独立的 δs 752
杜尔贝科(Dulbecco) 96
迪谢内肌营养不良(Duchemme's muscular dystrophy) 50
端化 699
端粒(telomere) 29,203,206
端粒的复制模型 290
端粒结合顺序(telomere-associated seqence) 206
端粒酶(telomerase) 288,592
短-补丁修复(short-patch repair) 657
短单拷贝序列(short single copy seguence, SSC) 621
短的分散序列(short interspersed sequence ,SINES) 222,757,760
短的头部片段(short headpiece) 456
短期调节(short-regulation) 530
短型 copia 756
短肢畸型(phcomelia) 60
段估计 21
断裂非重建性愈合 676
断裂基因(split gene) 227
断裂融合桥(Breakage fusion bridge) 667,735
断裂-愈合学说(breakage-rdunion hypothesis) 666
对氨甲酰基转移酶(trauscarbamylase) 537
对称分布 18
对称排列 454
对称型相互易位 697
多倍体(ployploid) 684,685
多蛋白(polyprotion) 746
多个着丝粒 530
多核苷酸磷酸化酶(polynucleotide phosphorylase) 364
多基因家族(multigene family) 850
多基因假设 98
多聚腺苷化(polyadenylation) 335
多拷贝抑制(multicopy inhibition) 732
多利绵羊 769
多联体(concatemer) 264
多瘤病毒(polyoma virus) 178
多纳休(Donahue) 96
多顺反子(polycistron) 322,449,771
多态座位比和杂合性(proportion of poly morphic loci and heterozygosity) 827
多肽转运装置(translocation machinery) 430
多体(multimers) 353,691
多体(polysome) 691
多萜醇(dolichol) 437
多线染色体(polytenne chromosomes) 531,681
多效现象(pleiotropism) 61
多样性基因(Diversity) 552
多指 10
多重跨膜结构域(multiple menbrane-spanning regions) 432
多重突变型(multiple auxotrophs) 137

额外的小染色体(extra small chroimosome) 210
额外环(extra arm) 377
二倍化(diploidization) 93
二倍体(diploid) 684
二次突变假设(two mutation hypothsis) 596
二级结构和翻译的调控 494
二价体(bivalent) 686
二粒小麦(Triticnm dicoccoide) 694
二联体(dyad) 30
二硫键 436
二氢叶酸还原酶(DHFR,dihydrofllate reductase) 423,534
二位点模型 415

法尼酯(farnesyl) 537
翻译(translation) 360
翻译调控 487
翻译功能区(translational domain) 386
翻译后转运(post-translation translocation) 420
反 Lepore 671
反接重复(reverse duplication) 670
反馈抑制(feedback inhibition) 444
反密码子臂(anticodon arm) 376
反式(anti) 176
反式(Trans) 77
反式 AB(trans AB) 65
反式剪接(trans-splicing) 351
反式显性(trans-dominant) 452
反式作用 543
反式作用 RNA(trans-acting RNA) 346
反式作用因子(trans-acting factor) 311,443
反向调节(retroregnlation) 292,527
反向遗传学(reverse genetics) 650
反向重复(inverted repeats) 219
反向重复序列(inverted repeats,IR) 621, 720,737
反选择(back-selected) 572
反义 RNA(antisense RNA) 495,586,732

反义 RNA 的调控　292,495
反义技术　497
反义链(antisen strand)　297
反应规范(norm of reaction)　58
反应停(thalidomide)　60
反应性元件(response elenents)　574
反终止作用(antitermination eefect)　507
反转录病毒(retroviruses)　744
反转录基因(retro gene)　220
反转录酶(reverse transcriptase)　744
反转录转座子(retrotransposons)　744
反转录子(retroposons)　744,753
泛蛋白连接酶体系(Uoiguitin ligase system)　590
泛生论(hypothesis of pangenesis)　1
泛子(pangens)　2
方差(variance)　103,834
纺锤体(spindle)　24
放射环和螺旋环共存模型　202
放射环模型(radeal loop model)　201
放射线　646
放线菌酮　620
非病毒超家族(nonviral suberfamily)　757
非对称型相互易位　676
非分泌型(nonsecrtors)　66
非复制型转座(nonreplicative transposition)　725,730
非贡献基因(noncontring alleles)　100
非孟德尔遗传(non-Mendelian inheritance)　623
非识别阅读框(unidentified reading frame,URF)　419
非同源性重组　671
非选择性标记(nonselected marker)　146
非整倍体(aneuploid)　684,691
非正常切离　163
非重建性愈合　666
非洲起源说　854
非洲爪蟾(Xenopus laevis)　768
非转录间隔区(NTS)　227
非自主发育(nonantonomous development)　116
非自主性因子(nonautonomous elements)　737
非组蛋白　195,565
肺小细胞癌(lung,small cell Ca)　610
费城染色体(Philadelphia chromosome,Ph)　612,692
分化(differentiate)　766
分节基因(segmentation genes)　776
分解代谢的阻遏(cata-bolite repression)　462
分界模型假设　794
分离律(Law of segregation)　9
分离型 pet-　628
分裂细胞核抗原(proliferating cell nucleus antigen PCNA)　280
分泌型(Secretor)　66
分母因素(denominator elements)　807
分散模型(Dispersive model)　234
分散因子(short interspersed elements, SINEs)　223
分生孢子(conidia)　83
分支点顺序(branch-point seguence)　345,346
分枝法(branching process)　17
分子伴侣(chaperone)　258,423,479
分子病(molecular diease)　128
分子寄生虫　636
分子驱动(molecular drive)　853
分子遗传学(molecular genetics)　816
分子因素(numerator elements)　807
分子杂交(molecular hybridization)　184
粉红眼基因 Car(Carnation)　78
粪壳菌(*Sodaria fimicola*)　702
封闭阅读框(block reading frame)　419
佛波酯(Phorbol ester)　608
弗轮克尔－康拉特(Heinz Fraenkel-Conrat)　167
辅阻遏物(corepressor)　444,449,471
负超螺旋　179,716
负的互补(negative complementation)　452
负调节(negative regulation)　446
负干涉(negative interference)　82
负控制　444
负链(minus strand)　353,494,746
负链强停止 DNA(minus strong-stop DNA)　748
负选型交配(negative assortative mating)　847
附加体(episome)　或游离基因 143,726
附着蛋白　493
附着基部(primary loop anchorage)　570
附着位点(attatchment sites, att)　161,714
复等位基因(multiple allelism)　64
复合寡聚糖(complex oligosacchrarides)　438
复合转座子(composite transposons)　722
复合座位(complex loci)　796
复性(renaturation)　180
复制不足(underreplication)　531
复制叉(replication fork)　238,258
复制叉陷阱　271
复制单位(replication unite)　286
复制的终点　239
复制酶　493
复制酶复合体　494
复制起始点　239
复制缺陷型(replication-defective)　751
复制体(replisome)　275
复制形(RF1)　262
复制型转座(replicative transposition)　725
复制原点(origin of replication)　190
复制原点控制起始的数量　532
复制子(replicon)　286

副节(parasegment) 788
副密码子(paracodon) 379
副染色体(accessory chromosome) 39
富兰克林(R. Franklin) 171

盖莫夫(G. Gamor) 361
概率(probability) 15
干扰 mRNA 的互补 RNA(mic-RNA, mRNA-interfering complementary RNA) 495
干涉(chromasomal interference) 82,708
甘露糖苷酶(ER mannosidase) 437
感染复数(multiplicity of infection, MOI) 526
感受态细胞(competent cells) 135
冈崎(Okazaki) 273
冈崎片段(Okazaki fragment) 273
高变区(hypervariable region) 551
高度重复顺序 223
高尔顿(Galton, F.) 2,107
高甘露糖寡聚糖(High mannose oligosaccharides) 437
高频重组(High frequence recombination,Hfr) 品系 143
高频转导(High-feequency transduction, HFT) 163
高亲和性位点 458
高泳动蛋白(High mobility group, HMG) 568,812
睾丸决定因子(人类:testis-determining factor,TDF; 小鼠:terstis-drtermining gene on the y, Tdy) 812
睾丸瘤(testicular tμmors) 610
格尔登(John Gurdon) 768
格拉博斯基(Grabowski, P. J.) 339
格赖德(Greider) 290
隔离子 569
根癌农杆菌(Agrobacterium tumefaciens) 193
工程转化(engineered transformation) 135
工业黑化现象 841
功能获得(gain-of-function) 786
功能区转换(domain-swapping) 579
供体(donor) 位点 135,338
供体元件 542
共翻译转运(co-translational translocation) 421
共合体(cointegrate) 698,724,729
共价闭合环状 DNA(cccDNA) 179
共价延伸 263
共上位(coepistatic) 69
共显性(codominance) 63
共转变(coconversion) 703
共转导(contransduc) 160
共转化(contransformation) 137
贡献等位基因(contributing alleles) 100
钩瓣属(*Clarkia*) 679
构巢曲霉菌(*Aspergillus nidulans*) 92
构象依赖(conformation dependent) 258

孤散子(orphons) 221
骨架蛋白(scaffolding protein) 570
骨架附着区(scaffold attached region,SAR) 570
固定杂种优势 673
寡聚物形成功能域(oligomerization domain) 392
光复活(photoreactivation) 656
光裂合酶(photolyase) 656
光修复(light repair) 656
广义遗传力(broad-sense heritability) 109
归巢(homing) 231,760
归巢内切酶(homing endonuclease) 440
鲑鱼 690
滚环复制(looped rolling circle replication) 260,264,531
果蝇性别控制基因 *da*(daughterless) 577
过量复制(coverreplication) 531
过氧化物酶体(peroxisomes) 426

哈迪-温伯格(Hardy-Weinberg)法则 820
哈迪-温伯格平衡(Hardy-Weinberg eguiliberum) 821
海斯(W. Hayes) 139
合胞体(syncytium) 529,772
合成前(presynthetic)错误控制 247
合子基因(zygotic genes) 788
合子诱导(zygotic induction) 155
核 mRNA 剪接 346
核定位信号(nuclear localization signals) 421
核苷(mucleotide) 169
核苷酸(nucleotides) 169
核苷酸结合折叠(nucleotide-binding fold) 392
核基质(nuclear matrix) 570
核基质结合区(matrix associatior region,MAR) 570
核酶(Ribozyme) 336
核内不均一 RNA(heterogenous nuclear RNA,hnRNA) 307,331
核仁组织区 531
核丝(nucleofilament) 200
核素(muclein) 164
核糖核蛋白(snRNP) 348
核糖体池 400
核糖体蛋白(r-proteins) 489
核糖体结合位点(ribosome-binding site) 399
核糖体亲和蛋白(Ribophorin) 430
核糖体依赖性 GTP 酶活性(ribosome-dependent GTPase activity) 404
核酮糖二磷酸羧化酶(ribulose bisphosphate decarboxylase 623
核外突变(extranuclear mutaation) 625
核纤层蛋白(lamin) 208
核小体(nucleosome)核粒 197,198
核小体核心颗粒(nucleosome core particle) 198
核小体模型 198
核心(core) 714

核心酶 250,301
核心启动子(core promoter) 315,573
核心区(core) 456
核心元件(core element) 315
核型(karyotype) 31,209
核型的改变 662
核质蛋白(nucleoplasmin) 282
黑化型(carbonaria) 841
黑麦(*Secale*) 689
黑尿酸氧化酶 115
黑色素瘤(melanoma) 610
黑色素体(melanosome) 122
黑酸尿病(alcaptonuria) 115
黑猩猩(*Pan troglodytes*) 680
黑爪异鳃(*Xenobius melanonychus*) 852
亨金(Henking,H) 39
恒定区(constant region,C) 551
红-黑田鼠(*Clethrionomys gapperi*) 825
红鲈鱼(*Sacura margaface*) 49
红霉素(erythromycin resistance) 630
红色链胞霉(*Neurospora Crassa*) 118
后部决定因子(Nanos) 782
后部系统(posterior system) 780
后期Ⅰ 30
后生蛉(Bonellia) 48
后随链 251,274
候雅明(Hou Ya-ming) 378
呼肠病毒(Reovirus) 334
琥珀(amber) 密码子 370,418
琥珀突变(amber) 369
互变异构移位(tautomeric shift) 644
互变异物体(tautomers) 643
互补效应(complement effect) 71
互斥事件(mutually exclusive events) 16
互加效应(adititive effect) 70
互连细胞(intercconnected cell) 774
花斑位置效应(variegated position effect) 672,677
华莱士(Russel Wallace) 840
滑动错配 653
滑面内质网(smooth ER) 427
化学校对(Chemical proofreading) 395
桦尺蠖(椒花蛾,*Biston betularia*) 841
还阳参属(*Crepis*) 669
环沟(ring canals) 774
环节动物 Ophryotrocha 804
环磷酸二酯酶(phosphodiesterase) 327
环形染色体 210
环状区域(looped domain) 201
环状染色体(ring chromosome) 593
黄嘌呤(xanthine,X) 648
黄曲霉毒素 B1(aflatosin B1,AFB1) 655
黄色霉素(Kirromycin) 413
黄鳝(Monopterus albus) 49
黄眼果蝇 634
恢复系:(N) Rf R 635
回复突变(reverse mutation) 641,831
回归(regressim)分析 107
回归线(regaression line) 108
回环模型 276,277
回交(back cross) 8
回文顺序 (Palindromic sequence) 219
活性暗箱(active cassette) 539,542
活性斑带 532
火山岛(Tristan da cwnha) 834
获得性免疫缺损缺合症(acuired immunodeficiency syndrome, azidothymidine,AIDS) 647
获救技术 782
霍尔丹定律(Haldane's law) 79
霍格内斯框(Hogness box) 310
霍利迪(Robin Holliday)701,704

几率(chance) 15
肌动蛋白基因(actin genes) 351
肌钙蛋白 T(troponin T) 230,583
肌红蛋白基因 850
肌球蛋白(myosin) 582
积量效应 70
基因(Gene) 10
基因 0.3 482
基因 0.7 482
基因 2 482
基因簇(gene cluster) 220,440
基因的精细作图 131
基因的扩增(amplification) 531
基因的转变(gene coversion) 439,699
基因家族(gene family) 220
基因间的 DNA(intergenic DNA) 190
基因库(gene pool) 816
基因流(gene flow) 836,839
基因内启动子(intragenenic promoter) 316
基因内抑制 362
基因频率(gene freguencie) 818
基因突变(game mutation) 640
基因外抑制 362
基因文库(gene bank 或 gene library) 185
基因型(genotype) 10
基因转变(gene conversion) 547
基因组突变(genome mutation) 640
基质(matrix) 425
基质(stroma) 425

基质导向信号(matrix-targeting signal) 424
激活物(activator) 444
激活因子(activator) 735
吉尔伯特(Gilbert. W) 228,263
级联效应(cascade effect) 781
极核(polar nuclei) 35
极化(Polarity) 705
极化子(polaron) 703
极细胞质 531
极限(crisis) 592
极性(polarity)突变 306
极性突变 719
极质(polar plasm) 777
急性淋巴白血病(ALL,acute lymphoblastic leukemia) 613
急性髓性白血病 693
脊髓灰质炎病毒(poliovirus) 268,410
剂量补偿(dosage conpensation) 564
剂量补偿效应 24-36) 811
剂量效应定位法 94
加工假基因(processed pseudogenes) 220,758
加帽 333
加尾位点 334
加尾位点的选择 582
加性方差(additiive genetic variance) 110
加字突变 362
家谱(Pedigree) 12
家系连锁分定位析法 96
家族性黑矇性痴呆(familial amaurotic idilcy) 125
甲基化(hemimethylated) 436,571
甲基化酶 498
甲基黄酸甲酯(MMS) 649
甲基黄酸乙酯(EMS) 649
甲酰化反应 622
假基因(pseudogene) 220
假连锁(pseudolinkage) 673,677
假型肥大(pseudomu-pertrophic) 50
间插序列(intervening sequence) 227
间隔序列 (spacer) 226
间期(interphase) 24
间性基因 ix(intersex) 809
兼性异染色质(facultatire heterochromatin) 33,195,563
减色效应(hpochromic effcef) 183
减数分裂(meiosis) 24,28
减数后分离(postmeiotic segreation) 702
减字突变 362
剪接前导 RNA(spliced leader RNA,SL RNA) 351
剪接体(spliceoome) 350
剪接体滞留模型(spliceosome retentior model) 580
剪接因子 348
简单端粒序列(simple telomeric sequence) 206
碱基(base) 169
碱基插入剂(inntercalatirg agents) 627
碱基对取代突变(base-pair substitution mutation) 640
碱基类似物 646
碱性 HLH(bHLH,proteim) 577
建立者效应(founder effect) 834
降钙素(calcitonin) 582
降解控制(degradation control) 580
降解质粒(degradative plasmids) 193
交叉(chiasmata 或 singular.chiasma) 30
交叉干涉(chiasma interference) 82
交叉理论 698
交叉型学说(chiasmatatype hypothesis) 698
交错切割 731
交互分离(alternate segregation) 677
交互重组 698
交换(cross-over) 30,77
交换(exchange) 666
交换复合体(crossover complex) 729
交换控制(crossver control) 709
交换值(crossing-over value) 79
铰连区 456
交配型(matig type) 83,537
交配型的转换 539,718
角化细胞(Keratinocyte) 122
绞花式遗传(crisscross inheri-tance) 50
校对(proofreading) 394
校正基因 379
校正控制(proofreading) 247
校正突变(suppressor mutation) 362,380
酵母(*S. cerevisiae*) 701
酵母半乳糖代谢 578
酵母的 2μ 环(two-micron circle) 635
酵母的交配型 537
酵母人工染色体(YAC) 203
酵母交配型转换 312
酵母小菌落(petife) 627
酵母转座子(transponson yeast ,Ty) 752
接合(conjugation) 135
接合管(conjucation tube) 141
节结(nodes) 707
结构纯合体(structral homozygote) 666
结构基因(structural genes) 446,447
结构杂合体(structral heterozygote) 666
截刚毛(bobbed) 53
姐妹染色单体(sister chromatids) 24
姐妹染色单体差别染色方法(sister-chromatid differential staining) 236,699
姐妹染色单体交换(sister chromatin exchage,SCE) 237
解离 729

解离酶(resolvase) 723,729
解离模型(直接作用) 455
解离序列(res) 723
解离因子(dissociator) 735
解链(melting) 180
解链温度或熔点(melting temperature, T_m) 180
解旋酶(helicase) 254
金(Thomas King) 767
金老梅(*Potentilla trutocosa*) 52
金砖(Goldbrick) 310
津得(N. Zinder) 158
紧密结合(tight binding) 301
近交(inbreeding) 847
进化速率 850
进位反应 412
进行性肌营养不良(progressive muscular dystorphy,DMD) 50
经典模型(classical model) 826
精母细胞悬液 683
精确切离(precise excision) 727
静息子(sisencer) 312
九聚体(nonamer) 555
居埃诺(Lucien Cuenot) 68
局部二倍体(merodoploidy) 146
局限性转导(restricted transduction) 161
聚合酶(polymerase) 296－304
决定(determination) 766
绝缘子(insulator) 569,570
均染色区(homogeneously staining region,HSR) 535
均值(Mean) 102
菌苔(lawn) 155

卡方,χ^2(Chi square method) 21
卡罗瑟斯(E. Carothers) 37
卡皮钦科(G. Karpechenko) 688
开放的启动子 301
开放二元复合体(open binary complex) 301
开放阅读框(open reading frame or ORF) 419
看家基因(Housekeeping genes) 442
抗癌基因(antioncogene) 592,594
抗癌基因 p53 592,595
抗缔合因子(antiassociation factors) 400
抗副中肾激素(antitimullerian hormone, AMH) 811,813
抗红霉素基因 498
抗维生素 D 佝偻病(Vitamin Dresistant rickets) 51
抗性质粒(Resistance (R) plasmids) 193
抗性转移因子(resistance-transfer factor, RTF) 146
抗叶锈病基因 694
抗胰蛋白酶核心(trypsin-resistant core) 455
抗终止蛋白(antitermination protein) 508
抗终止因子 444
拷贝选择(copy choice) 749
科恩伯格(A. Kornberg) 245
科拉纳(Khorara) 365
科伦斯(Carl Corrans) 2, 619
可变区(variable region, V) 551
可变数同向重复序列(variable number tandem repeat VNTRs)区 224
可诱导(inducible) 444
可阻遏(repressible) 444
克拉花(C. biloba) 679
克莱顿(Creighton, B) 77
克兰费尔特综合征 (Klinefelter's syndrome) 55
克里克 (F. H. C Crick.) 292,373,375
克隆(cloning) 767
克隆嵌板法(clone panel method) 95
克隆选择学说 552
克卢格(Klug) 198
空载 tRNA 499
空转反应(idiling reaction) 500
孔道途径(pore pathway) 435
控制位点(controlling site) 443
控制因子(controlling elements) 735
枯草杆菌(*Bacillus subtilis*) 135,475
跨膜结构域(transmembrane domain) 431
快停突变(fast-stop mutants) 245
扩增 534

拉链作用(Zippering) 180
莱昂(Mary Lyon) 564
莱昂化(lyonization) 564
莱昂假说(Lyon hypothesiis) 564
莱德伯格(E. Lederberg) 153
莱德伯格(J. Lederberg) 137,158
莱－尼二氏症(Lesch-Nyhan syndrome) 123
劳伦斯肌肉(musele of Laurence, MOL) 810
劳斯(P. Rous) 594
劳斯肉瘤病毒 (Rous sarcoma virus) 594
雷迪(E. P. Reddy) 606
类别转换(Class switching) 560
类病毒(viroids) 353
类插入序列(IS-like elements) 722
类核(nucleoid) 178,188
类核区(nuclecid regions) 616
类囊膜(thylakoid membrane) 425
冷链(cord strand) 238
厘摩(centimorgan cM) 80
里奇 (A. Rich) 176
理想群体 821
利根川进(Tonegawa, S.) 552
利什曼原虫(*Leishmania*) 357

蛎壳翅(beaded ,Bd) 675
连接 DNA (linker DNA) 198
连接酶(ligase) 251
连接(Joining)区 J 552
连锁法则(law of linkage) 76
连锁群(Linkage group) 79
联会(synapsed) 29
联会复合体(synaptonemal complex, SC) 29,683,707
镰刀形细胞特征(sickle cell trait) 128
镰形细胞贫血(sickle cell anenia) 63,128
链孢霉(*Neurospora crassa*) 83
链的置换 249
链环数或称拓扑环绕数, L(Linking number) 179
链霉素(streptomycin) 620
链霉素的抗性(smr) 630
链霉素敏感型(sms) 630
链终止密码子(chain-terminating codons) 372
两相性(diphasic) 484
两性基因 *her*(hermaphrodite) 809
两性畸形(gynandromorphs) 692
两性嵌合体 692
两性体(hermaphrodites) 772
亮氨酸拉链(Leucine zipper) 577,609
烈性(virulent)突变 509
烈性噬菌体(virulent) 152
烈性突变 513,521
裂解蛋白 493
裂解途径(lytic pathway) 152
裂解周期 504, 513
裂口(Clefts) 384
裂缺(Gap) 207
裂缺基因(gap) 789
裂手裂足 10
邻近分离(adjacent segregation) 677
磷酸化 436
磷酸化酶 589
磷酸基团(phosphate group) 169
零体 691
绿白斑(albomaculata) 623
氯霉素 (chloromycetin) 620
孪生斑(twin spots) 91
卵巢乳突状癌(ovarian papillary Ca) 612
卵睾(ovotestes) 55,813
卵黄蛋白原(vetellogenin) 571
卵壳蛋白 532
卵壳基因 532
卵细胞(eggcell or ovom) 35
卵周隙(perivitelline space) 785
罗伯特(R.J. Roberts) 227
罗伯逊易位(Robertsonian transloca－tion) 676
萝卜甘蓝(*Raphanobrassice*) 688
螺旋管(solenoid)模型 200
螺旋环模型 201
螺旋－环－螺旋(HLH, helix-loop-helix) 577,807
螺旋折叠(helically folded) 202
螺旋－转角－螺旋(helix-turn-helix, HTH) 456,517,574

麻疹副粘病毒(*measles paramyxoviruses*) 357
马蛔虫(*Parascaris equoorum*) 530
马默多蒂(Marmur-Doty)关系式 181
麦克库西克(Mckusick) 124
麦克朗(McClung, C.E.) 39
麦克林托克(McClintock) 77,206,718,735
麦芽糖操纵子 502
脉冲标记实验(pulse-labeling experiment) 273
曼岛猫(Manx cat) 68
曼陀罗(*Datura stramonium*) 37
慢停突变(slow-stop mutants) 245
慢性髓性白血病(CML, chronic myelogenous leukemia) 613, 693
猫叫综合征(cri-du-chat syndrome,或 cat cry syndrome) 210, 668,664
毛地黄(*Digitalis purpuread*) 71
毛耳(hairy ears) 52
锚(anchor) 431
帽结合蛋白(CBP) 410
帽子(cap) 334
梅塞尔森－拉丁(Meselson-Raldding)重组模型 704
梅塞尔森－斯塔尔(Meselson-Stahl)实验 234
酶-启动子二元复合物(closed binary complex) 299,301
孟德尔(G. Mendel) 2,7
孟德尔第二定律 13
孟德尔式群体(Mendelian population) 816
孟买型 65
米斯切尔(J.T. Miescher) 164
米组诺(T.Mizuno) 552
密度制约抑制(density-dependent inhibition) 593
免疫区(immuinty region) 514
免疫球蛋白重链基因重排易位 612
面对面核苷酸结构(confrontde nucleotide structure) 334
命运(fates) 766
命运图 788
模板选择(copy choice)性重组 698
模板选择学说(copy choice) 700
模板转换(template-switching) 249
膜导向信号(membrane-targeting sigmal) 424
膜结合多体(membrane-bound polysomes) 421
摩尔根(T.H. Morgan) 40
魔斑Ⅰ 499
魔斑Ⅱ 499

末端蛋白(TP)　268
末端区域(terminal region)　438
末端缺失(terminal deletion)　667
末端重复序列(long terminal repeat ,LTR)　749
末端系统(terminal system)　781
末期 Ⅰ　30
莫诺 (J. Monod).　295,443
母体基因(maternal genes)　776
母体生殖细胞基因(maternal germline genes)　779
母体体细胞基因(maternal somatic genes)　779
母体效应基因　807
母体遗传(maternal inheritance)　623
母体影响(maternal effect)　633
母细胞(blast cells)　772
木村资生　828
穆勒(H. J. Muller)　206,670

内部蛋白　481
内部核糖体进入位点(internal ribosome entry site, IERS)　410
内部核心(inner core)　438
内部控制区(internal contron regin ,ICR)　316
内部顺式控制区　724
内部引导顺序(internal guide seguence'IGS)　341
内部终止子(intrinsic terminators)　304,510
内部转录间隔序列 (ITS)　329
内蛋白子(intein)　439
内蛋白子归巢(intein homing)　439,761
内含子(intron)　227－230
内含子归巢(intron homing)　756
内基因子(endogenote)　149
内膜(inner membrane)　423,433
内生病毒(endogenous provirus)　744
内式(endo)　176
内-糖苷酶 H(endo-glycosidase H, Endo H)　438
内源前病毒(endogenous provirus)　750
内源序列　534
纳精囊　805
奈特(T. Knight)　2
南美仓鼠(*Akodone*)813
囊胚(blastula)　768
囊胚形成　808
尼尔松－埃勒(Nilsson-Ehle)　98
尼伦伯格(Nirenberg. M. W)　364
拟病毒(virusoids)　353
拟等位基因(pseudoallele)　130
拟谷盗属(*Tribolium*)　796
拟显性(pseudo-dominante)　668
酿酒酵母(*S. cerevisial*)　588
鸟苷释放因子(GNRF, guanine nucleotide relase factor)　606
鸟苷转移酶　333

尿嘧啶 N-糖苷酶(uracil N-glycosylase)　273
扭曲数 W(Writhing number)　180
扭转角　177
钮体　197
浓缩期(diakinesis)　30
女娄菜(*Melandrium album*)　52
诺尔(M. Noll)　197
Northern 杂交(Northern hybridization)　185

欧洲鳗鲡(Anguilla anguilla)　49
偶线期(zygotene)　25

拌环模型 202
庞尼特(Punnet, R. C)　75
胚囊(embryo sac)　35
胚乳(endosperm)　35
胚乳母细胞(endosperm mother cell)　35
胚系(germ line)　模式　549
胚系学说　552
配对规则基因(pair-rule)　789
配子体(gemetophype)　35
喷瓜(*Ecballium elaterium*)　47
嘌呤霉素(puromycin)　387,414
频率(frequency)　15
平衡模型(balance model)　455,826
平衡选择　827
平衡易位　677
平衡致死品系(balanced lethal system)　675
平截　641
平均适应值　843
平均值(average)　102
平行进化同源基因(paralogs)　802
平行连接(Pra-nernic joint)　712
瓶颈效应(bottleneck dffect)　836
珀尔(Perler)　439
葡萄球菌核酸酶(Staphylococcal nuclease)　197
葡萄糖苷酶(glucosidases Ⅰ和Ⅱ)　437
葡萄糖磷酸变位酶(PGM)　819
普遍性途径(general pathway)　433,435
普遍性重组(generalized recombination)　698
普遍性转导(generalized transduction)　158
普里布诺盒(Pribnous box)　299
普通小麦(*T. aestivum*)　689,695

七聚体(heptamer)　555
七鳃鳗　852
棋盘法(Punnett square)　17
启动子(promoter)　298
启动子 P_{IN}　732
启动子 P_{OUT}　732

起点(start) 587
起始(initiate) 594
起始阶段(START stage) 549
起始密码子(initiation codon) 370,401
起始密码子重叠 487
起始因子(initiation factors,IF) 399
起始因子 IF-1 400
起始因子 IF-2 400
起始因子 IF-3 400
起始子(initiator, Inr.) 309,402
扦插(cuttings) 767
迁移(migration) 839
前孢子(forespore) 476
前病毒 DNA(proviral DNA) 744
前部系统(anterior system) 780
前导链(leading strand) 251,258,274
前导顺序(leader) 408,410
前导肽(Leader sequence) 422
前导序列 (leader sequence) 322
前导序列(trpL) 470
前后轴(anterior-posterior) 776
前期(prophase) 24,28
前起始复合体(preinitiation complex) 251,319
前突变(premutation) 653
前胰岛素原(preproinsulin) 439
前引发复合体(prepriming complex) 258
前引发体(preprimosome) 261
钱宝(Chambon)法则 338
钱宝(P.Chambon) 197,227
钱普(S.Champe) 368
嵌合体(genetic mosaic) 692
嵌镶上位(mosaic epistatic) 69
嵌镶显性(mosaic dominance) 63
腔(lumen) 425
强直性肌营养不良(Myotonic dystrophy) 654
强终止子 306
羟基化 436
羟基磷灰石 183
桥梁杂交(bridging cross) 694
切除修复(excixion-repair) 656
切尔迈克(Tschermak) 2
切赫(T.Cech) 336,358
切口平移(nick translation) 249
切离 714
亲核进攻 339
亲组合(parental combination) 13
侵入性质粒(virulence plasmids) 193
轻度重复顺序 219
轻链(Light chaim) 242,551
清晰的噬菌斑(Clear plaquea) 513
秋水仙素 688
取代环(displaced loop) 242
取样误差(sampling error) 833
去甲酰化酶(deformylase) 404
全保留复制模型(conservetive model) 234
全程起始(de novo initiation) 262
缺刻(nick) 207
缺刻翅(notch) 669
缺失(deletion 或 deficiency) 652,666,667
缺体(nullisomy) 691
确立细胞株(Establised cell line) 593
群体遗传学(population genetics) 816

染色单体(chromatids) 24,30,705,707
染色单体转变(chromatid conversion) 705
染色体丢失 530
染色体骨架(scaffold) 201
染色体畸变(chromosomal aberration) 640,665
染色体畸变定位法 95
染色体加倍 688
染色体间易位(interchromosomal shift) 676
染色体间重组(interchromosomal recombination) 77
染色体内移位(intrachromosomal shift) 676
染色体内重组(intrachromosomal recombination) 77
染色体凝聚调节蛋白(RCC1 ,regulator of chromosome comdensation) 591
染色体桥(chromosomal bridge) 203,667
染色体突变(chromosome mutation) 640,665
染色体中心(chromocenaer) 531
染色体重排(chromosomal rearrangements) 666
染色体组(chromoxome sets) 665
染色体组分析 679
染色质(chromatin) 195
热点选择 720
热激蛋白(heat shock protein) 423
热链(hot strand) 238
热激基因 hsp(heat shock genes) 479,569
人类(*Homo sapiens*) 680
人类基因组计划 194
人类免疫缺陷病毒-I(human immunodeficiency rirus-1,HIV-1) 647
绒毛膜细胞 125
绒泡菌(*Physarum polycephalum*) 339
溶素原(lyxogens) 513
溶原(lysogeny) 152
溶原化周期 504
溶原途径(lysogenic pathway) 152
溶原性细菌(lysogeny bacteria) 152
融合基因(fusion gere) 423,613
融合内蛋白子(fused intein) 439

融合桥断裂循环(fusion-bridge-breakage cycle) 737
乳酸脱氢酶同工酶 673,852
弱终止子 304

萨瑟杂交(Southern blotting) 185,567
塞克斯坦姆盒(Sextama box) 300
塞勒(Seiler,J) 39
腮瘤 612
三倍体(triploid) (3x) 684,685
三点测交(three - point test cross) 80
三价体(trivalents) 686
三联体结合实验 365
三色猫 564
三体(trisomy) 691
三位点模型 415
三系二区制种法 635
三叶草型(cloverleaf) 376
三元复合体(ternary complex) 301,407
三中读二(2 out of 3 reading) 373
伞花山羊草(*Aegilops umbellulata*) 694
伞毛蛋白(pilin) 141
扫描(scaning) 模型 411
色氨酸操纵子 469
色盲等位基因的频率 825
色素(clolor) 61
杀伤性酵母 632
沙门氏菌 (s. typhimrium) 484,651,718
上升突变(up mutation) 299
上位基因(epistatic genes) 68
上游激活序列(upstream activating seguences UASs) 312
上游控制元件(upstream,control element UCE) 315
上游启动子成分 573
身分(identity) 378
深红虎蛾(*Panaxia dominula*) 817
神经母细胞瘤(Neuroblastoma) 603,610
神经肽(neuropeptide) 582
渗震作用 167
绳珠模型 198
失败-保险的终止位点 319
十六烷酸 606
十四烷基化位点 602,604
十字配对 677
石斑鱼(*Ophrgtrocha Puerilis*) 49
时序调控 Lin-14 773
时序调控基因(chronogenes) 773
识别螺旋(recognition helix) 517
视网膜母细胞瘤(Rb,Retinoblastoma) 68,592
适合度(fitnees) 842
释放因子 eRF(release factor) 372,418
噬菌斑(plaque) 155
噬菌体 MS2,R17,f2 和 Qβ 493
受纳群体 839
受体(acceptor)位点 338,542
受体(recipient) 135
受体臂(acceptor arm) 376
受体—反受体之间(receptor-counter receptor) 相互作用 784
受体转录因子(receptor transcription,PRTF) 546
束缚因子(commitment factor) 313
衰减模型 471
衰减子(attenuator) 470
双翅目昆虫 681
双二倍体(amphidiploid) 688
双精受精开米拉(dispermic chimera) 55
双链断裂 731
双链断裂起始重组模型 705
双螺旋模型(double helix model) 172
双螺旋稳定蛋白(helix destabilizing protein) 254
双亲遗传(biparental inheritance) 630
双氢尿嘧啶环 376
双三体(double trisomy) 691
双受精(double fertilization) 35
双胎开米拉(twin chimera) 55
双体(disomy) 691
双微体(double-minute chromosomes) 535
双显性上位(double dominant epistasis) 70
双线期(diplotene) 30
双向电泳法 240
双胸复合体(bithorax comples , BX-C) 798
双隐性上位(double recessive epistatic) 71
双重倒位(duplicative inversions) 727
双重启动子(double promoter) 325
双重溶原(double lysoge) 163
双重性别基因 dsx(doublesex) 805
双着丝点染色体 210,667
水稻(*Oryza sativa*) 617
水蜡虫(*Pseudococcus nipae*) 563
水毛茛(*Ranunculus aguatilis*) 57
顺反测验(cis-trans test)131
顺反子(cistrons) 131
顺接重复(tandem duplication) 670
顺式(cis) 77
顺式(syn) 176
顺式 A. B. (cis AB) 65
顺式剪接(cis-splicing) 351
顺式-显性(cis-dominant) 451
顺式优先(cis-Preence) 733
顺式作用序列(cis-acting seguence) 311
顺式作用元件(cis-acting element) 443
顺向重复 219
顺序四分子(ordered tetrad) 84

斯特蒂文特(Sturtevant, A. H) 80,115
斯特恩(Stern, C.) 53,78
死亡之门(Kalilo) 636
四倍体(tetraploie) (4x) 684
四倍体番茄 687
四倍体曼陀罗(Datura stramonium) 687
四分子分析(terard analysis) 84
四价体(guadriva-lents) 686
四联体(tetrad) 30
四密码子家族盒(rour-codon family box) 617
四膜虫(*Tetrahymeno thermophila*) 195,206,338, 371,618
四型(tetratyne) 86
松弛型(relaxed)质粒 194
松驰型 DNA (relax form) 178
松驰型突变(relaxed(rel)mutants) 500
宿主范围 674
宿主基因 hfLA 526
宿主基因 hfLB(hfL, high frequence lysogenization) 526
宿主整合因子(integration host factor, IHF) 714
随机交配(random mating) 821
随机选择 720
梭链孢酸(frsidic acid) 417
羧基末端功能区(carboxy-terminal domain CTD) 308
缩合(condensation) 333
索斯泰克(Szostak, J. W.) 705

塔特姆 (Tatum) 115,137
胎儿镜(fetoscope) 125
肽基转移酶位点 386,387
肽转移酶(prptityl transferase) 414
泰宾(Tabin C. J.) 602
泰勒(Taylor)实验 236
探针(probe) 184
唐(Langdon Down) 691
唐尼(Downey)模型 280
唐氏综合征(Down's syndrone) 210,677,691
糖基化 436
糖基转移酶(goycosyl transferase enzyme) 437
套索(lariat)结构 345
特化染色质结构(scs 和 scs' ; specislized chromatim structures) 569
特纳氏综合征(Turner's syndrome) 55
特殊细胞中的启动子成分 573
特殊重组 698
特异性转导(specialized transduction) 158,161
剔除(knockout) 814
梯敏(H. Temin) 293,594
体节(segment) 788
体节极性基因(segment polarity) 789
体细胞突变(somatic mutation) 640
体细胞杂交定位法 95
体细胞重组(somatic recobination) 学说 552
体液免疫 550
体质(somatoplasm) 2
天才遗传(Hereditary genius 2
条件概率(conditional probability) 16
调节基因(regulator genes) 443
调节型基因(regulated genes) 442
调节子(regulon) 468
跳格 653
跳跃基因 718
贴壁依赖性(Anchorage dependence) 953
停止转运(stop-transfer) 431
通读(readthrough) 305
通用反式作用因子 574
通用性简并性(degeneracy) 372
同臂重复 670
同步培养 240
同二聚体(honwdimers) 577
同工受体 tRNA(isoaceptor tRNAs) 849
同裂酶(isoschizomers) 571
同卵双生子(identical monozygotic twins) 60
同配性别(homogametic sex) 42
同形(homomorphic) 性染色体 45
同型半胱氨酸内脂(homocysteine thiolactone) 395
同义密码子(synonyms) 372
同源多倍体(autopolyploids) 685
同源结构域(homrodomain) 801
同源染色体(homologous chromosome) 29
同源特异重组 698,718
同源异型蛋白(homeotic proteins) 575
同源异型盒(homeobox) 575,577,801
同源异型基因(homeoticgenes) 776,795
同源异型基因复合体(homeotic-complex, HOM-C) 801
同源异型结构域(Homeodomains, HD) 577
同源异型座位(homeotic loci) 577
同源重组 (homologous recombination) 696,698
同宗配合(homothallic) 139
同族 tRNAs(cognate tRNAs) 390
酮式 646
突变(mutation) 640
突变率(mutation rate) 642
突变频率(mutation frequency) 642
突变热点(mutational hot spots) 645
图距单位 m. u. (map unit) 80
退火(annealing) 180
拖尾顺序(trailer) 410
脱氨(基) (deaminaation) 645
脱甲酰酶(deformylase) 435
脱嘌呤(depurination) 645

脱酰 tRNA(deerylated tDNA)　395
驼背基因(hunchback)　780

外部的转录间隔序列(ETS)　329
外部引导理论　322
外蛋白子(estein)　439
外基因子(exogenote)　149
外激素(pheromones)　537
外膜(outer membrane)　375,433
外膜蛋白 OmpC 和 OmpF　495
外式 (exo)　176
外显子 (exon)　227
外遗传的(epigenetic)　738
完全连锁 (complete linkage)　79
烷化剂　649
烷基转移酶(Alkyltransferases)　656
晚期基因　504
晚期启动子(PE)　808
晚期转录　482
晚早期　504
晚早期基因(delayed early genes)　508
王惠君(Wang A.H.J)　176
网织红细胞　585
威尔金斯(Wilkins,M)　171
威尔姆氏(Wilm's)　瘤　610
威尔逊(Wilson, E.B)　39
微铺展(microspeading)法　683
微球菌核酸酶(micrococcal nuclease)　198
微小病毒(parvovirus)　269
微小染色体　535
韦格尔(Jeam Weigle)　663
维尔默特(Ian Wilmut)　769
卫星 RNAs(satellite RNAs)　353
卫生 DNA　223
末端系统(terminal system)　781
位点特异性重组(site-specific recombination)　698
位置效应(position effect)　671
位置信息　776
温和性噬菌体(temperate)　152
文昌鱼(*Amphioxus*)　802
稳定位置效应(stable position effect)　672
沃尔曼(Wollman)　147
无 Pu 和 Py 位点(apurinic or apyrimidinic ,AP)　273
无汗性外胚层发育不良(anhidrotic ectodermal dysplasia)　564
无女儿 *da*(daughterless)　806
无细胞系统　364
无限生长(Immortalization)　592
无效起始(abortive initiation)　301
无效重排(nonproductive rearrangement)　559
无序四分子分析(unordered tetrad analysis)　89
无义密码子(nonsense codons)　372
无义突变(nonsense mutation)　305,418,641
无义抑制(nonsense suppressor)　380
舞蹈病(Huntington's disease)　10
戊糖(pentose)　169
物理图谱　185

希布纳(Huebner,R.I)　594
希克斯(Hicks,T.B.)　539
烯醇式　646
稀有密码子　497
细胞癌基因(cellular orogene,c-onc)　594
细胞蛋白 mdm 2　601
细胞骨架的形成(cytoskeletal organization)　593
细胞免疫　550
细胞色素 B 基因(CoB)　339,358
细胞型(cytotyope)　743
细胞学定位法　96
细胞质分离和重组(cytoplasmic segregation and recombinatim, CSAR)　624
细胞质突变(cytoplasmic mutation)　625
细胞质雄性不育　560,635
细胞周期(cell cycle)　24,550,587
细胞周期调控　587
细线期(leptotene)　25
狭义遗传力(narrow-sense heritability)　109
下降突变(down mutation)　299
下位基因(hupostatic genes)　68
下游启动子(downwtream promoter)　316
夏皮罗(Shapiro)　718,729
夏普(Sharp ,P.A.)　227
先成熟　329
先天性成骨不全(Osteogenesis inperfecta)　61
先天性代谢酶缺陷　115
先天性睾丸发育不全　55
先天性卵巢发育不全　55
先天性性腺发育不全　55
先天愚型　691
先证者(proband or propositus)　12
纤毛虫(单细胞的原生动物)　371
显性(dominate)　8
显性方差(dominance variance)　110
显性上位(dominant epistasis)　71
显性失活(dominant negatives)　452
显性致死(dioninat lethal)　68
线虫(*C. elagan*)　352,766,771
线粒体 DNA(mtDNA)　615
线粒体的脑脊髓病(mitochondlrial encephalomyopathies)　652
线粒体核糖体　619
线粒体质粒　635

线粒体质粒 Kal DNA 636
线状质粒 S1 和 S2 635
限定子(delimiter) 570
限制性酶的精确分析法 94
限制性片段长度多态性(Restriction Fragment Length Polymorphism,RFLPs) 185,829
腺病毒(adenovirus) 268
腺病毒的癌蛋白 E1B 602
腺苷环化酶(adenylate cyclase) 462
相乘定律 15
相斥(repulsion) 76
相的改变(changes of phase) 739
相关(correlation) 105
相关系数(correlation coefficient) 105
相互易位(reciprocal translocation) 676
相加定律 16
相引(coupling) 76
相转变(phase variation) 484,718
香豌豆(*Lathyrus doratus*) 75
硝酸银染色 606
小 t 蛋白 229
小孢子发生(microsporogenesis) 35
小孢子母细胞(microspore mother cell) 35
小麦瘿蚊(*Mayetiole destructor*) 530
小分子胞质内 RNA(small cytoplasmic RNAs,ScRNA) 348
小分子核内 RNA(small nuclear RNAs, snRNA) 348
小睾症 55
小沟(minor or narrow groove) 172
小管网 427
小黑麦(Triticale) 688
小菌落(poky) 625
小卫星(minisatellite) 224
小型染色体(microchromosome) 32
小熊猫(Ailurus julgens) 218
效应区(effector functions) 551
效应物(effectors) 443
协方差(covariance) 105
协同调控(coordinate regulation) 449
协同进化(concerted evolution) 853
协同效应 520
携带者(carrier) 12,50,125,677,679
辛格(B. Singre) 167
辛普森(L. Simpson) 358
锌指(zinc finger) 575
锌指蛋白(Zinc finger protein) 317
新莱昂(Lyon)假说 45,564
新霉素(neomycin 或 neamine) 620
新生蛋白(nascent protein) 427
新生儿溶血症 66
新生霉素(novobiocin) 244
信号传递(signal transduction) 539
信号端(signal ends) 556
信号假说(signal hypothesis) 427
信号识别蛋白(signal recognition particle ,SRP) 428
信号肽(signal sequence) 426
信号肽酶(signal peptidase) 430
信使 RNA(Messenger RNA) 295
信息素(pheromore) 546
猩猩(Pongo pygmaeus) 680
形成孢子 539
形成素(morphogen) 780
性别分化(sex differentiation) 44
性别决定(sex determination) 44
性别决定基因 45
性别致死 Sxl(sex-lethal) 805
性别转换基因 tra(transformer) 49,805
性导(sexduction) 145
性反转(sex reversal) 49,812
性嵴 812
性染色体(sex chromosome) 33
性伞毛 141
性梳(sex comb) 810
性限制遗传(sex-limtited inheritance) 55
性相关遗传(sex-related inheritance) 49
性指数(sex index) 46,805
雄性不育系(s) rfr 635
雄性特异致死效应基因(male-specific lethal effects) 811
雄性同配性别 46
雄性途径基因 (MPGs) 815
熊猫起源 218,681
秀丽隐杆线虫(*C. elegans*) 352,766,771
溴化3,8－二氮基－5－乙基－6－苯基菲(etnidium bramide) 649
溴化乙锭(EB) 627
旋转酶(Gyrase) 256
选择标记(selected marker) 146
选择不同外显子 583
选择系数(selection coefficient) 842
选择子(selector) 310
血红素(neme) 128
血清(或生长因子) 依赖性(serum or growth facto dependence) 590,593
血清紧张素(5-羟色胺) (serotonin) 113
血小板生长因子(PDGF) 604
血友病(hemophilia) 50

芽盘(imaginal discs) 116,766
雅各布(F. Jacob) 149,295,443
亚基因组的(subgenmic) 信使 746
亚诺夫斯基(yanofsky) 367,471

亚硝基胍(NG) 649
亚硝酸(introus acid, NA) 648
烟草 621
延宕(ribosome stalling) 472
延伸因子(elongation factor) 412
严紧反应(strigent response) 499
严紧型(stringent)质粒 194
严紧因子(stringent factor) 500
羊膜穿刺 125
羊膜镜或宫腔镜 125
羊头原鲷 49
氧化脱氨损伤碱基 648
氧化作用(oxidatively damaged bases) 645
样本的机误(chance errors) 832
姚(Yao M. C) 290
摇蚊(*Chironomus*) 681
野生型[rho+N] 628
叶绿体 DNA(cpDNA) 621
叶绿体核糖体 622
一倍体数(x)(monoploid number(x)) 684
一基因一酶(one gene-one enyme)学说 118
伊甸园 DNA(Garden of Eden DNA) 223
伊斯特(E. East) 100
衣藻(*chlamydomonas*) 351,621,629
胰岛素原(proinsulin) 439
移框(frameshift) 356
移码突变(frameshift mutation) 641
移位(shift translocation) 676
移位(translocation)反应 397,415
遗传力(heritability) 108
遗传密码 360
遗传漂变(genetic drift) 821,832
遗传性代谢病(inherited metabolic disease) 120
遗传性酶病(hereditary enzymopathy) 120
遗传咨询(Genetic counseling) 124
遗传作图(genetic mapping) 80
乙基化 436
异臂重复 670
异丙基-β-D-硫代半乳糖苷(isopropylthiogalactoside, IPTG) 449
异常的端着丝点染色体 210
异常分离 702
异二聚体(heferedimers) 577
异核体(heterckaryon) 93,627
异核体测验 627
异卵双生(fraternal dizygotic tioins) 60
异配性别(heterogametic sex) 42
异染色质(heterochromatin) 33, 195,563
异时基因(heterochronic) 767
异位表达(ectopic expression) 780
异位重复(displaced duplication) 670
异形(heterochromatic)性染色体 45
异形配对(heteromorphic pair) 37
异源多倍体(allopolyploids) 685,687
异源嵌合体(chimera) 692
异源双链定位法(Heferoduplex mapping) 187
异宗配合(heterothallic) 630
抑制基因(inhibitor) 72
抑制基因(suppressor) 73,379
抑制突变(suppressor mutation) 641
抑制型小菌落[P-S] 628
易感性(snsceptiblilty) 438
易位(translocation) 612,666,676
引发酶(primase) 258
引物模板 251
引发体(primosome) 256
引发体结合位点(primosome assembly sity, pas) 262
引物结合序列(primer binding seguence, PB) 748
隐蔽卫星 DNA(cryptic satellite) DNA 223
隐匿基因 357
隐性(recessive) 8
隐性上位(recessive epistasis) 70
隐性致死(recessive lethal) 68
应激蛋白(stress protein) 258,423
英格拉姆(V. M. Ingram) 128
营养期(vegetative phase) 476
营养缺陷型(auxotroph) 84
优势比率(potence raaton) 112
游离内蛋白子(free intein) 440
游仆虫(*Euplotes octacarinatus*) 372
友伴双胎 55
有丝分裂(mitosis) 23,24
有丝分裂分离 624
有丝分裂重组(mitotic recombination) 91
有效种群大小(effective population size) 833
有效重排(productive rearrargement) 559
有义链(sense strand) 297
有义密码子(sense codons) 372
幼虫唾腺 681
诱变剂(mutagen) 640
诱导(induction) 444,448
诱导物(inducers) 443,449
诱发突变(induced mutation) 640
玉米转座因子 737
预杂交 185
原癌基因(proto-oncogene) 594
原肠胚(gastrula) 768
原代细胞(Primary cells) 593
原黄素(proflavin) 362,649
原生殖细胞(primardial germ cell) 34

原始的α-珠蛋白基因 851
原始的β-珠蛋白基因 851
原噬菌体(prophage) 152,513,714
原位(*in situ*)序列 443
原位分子杂交(*in situ* hybridization) 96,184
原位杂交定位法 96
原养型(prototroph) 84
远交(outbreeding) 847
远上游序列(far upstream seguence) 311, 573
约翰逊(Johannsen) 10
可读框(reading frame) 363
芸苔属(*Brassica*) 689
运输控制(transport control) 583

杂合优势(heterozygote advemtage) 847
杂合子(heterozygote 10
杂种 RNA 结构 353
杂种不育(hybrid dysgenesis) 740
甾类抗生素(steroid antibiotic) 417
载脂蛋白-B(apolipoprotein-B apo-B)基因 355
再程序化(reprogramming) 770
早期启动子(PL) 808
早秃(Baldness) 53
早早期 504
早早期基因(immediate early genes) 508
增强子(enhancer) 310
增色效应(hyperchromic effect) 180
折迭区域(folding domains) 232
折叠 436
赭石(ochre)密码子 369,418
赭石突变 369,418
蔗糖酶 581
真两性畸型 55
整倍体(enpeoid) 684
整臂融合(whole-arm fusion) 676
整合(intergrated) 714
整合酶(integrase) 744
整合膜蛋白(integral membrane protein) 422
整合宿主因子(IHF:integration host factor) 190
整合体(intasome) 716
整体开米拉(whole body chimerism) 55
正常传递男性(normally transmitting males,NTMs) 653
正常的切离 163
正常染色体(A-chromosome, A-染色体) 32
正超螺旋 179
正调控 444
正反交(reciprocal cross) 8
正干涉(positive interference) 82
正链(plus strand) 353,494
正链 DNA(plus strand DNA) 746
正链病毒(plus strand viruses) 746
正向突变(forward mutation) 641,831
正向重复(direct repeats) 219,720
正选型交配(Positive assortative mating) 847
支持细胞(或称足细胞,Sertoli's cells) 814
支架 (scafford) 188
枝原体(*Mycoplasma capricolum*) 371
脂双层(hydrophobic lipid bilayer) 422
直接修复(Direct repair) 655
植物的豆血红蛋白基因 851
指导 RNA(gRNA) 357
指甲-髌骨综合征(Nail-petalla syndrome) 61
指令学说 552
治愈(curing) 154
质粒(plasmid) 141,193
质粒 ColE1 292
质粒 PCM 255
致癌剂(carcinogens) 651
致癌物(carcinogenic) 594
致死 668
致死等位基因(lethal allele) 68
致育因子(Fertility (F)plasmids) 193
置换(exchange) 709
置信界限(confidence limits) 834
中部核心结构(Central core strucature) 341
中度重复顺序 219
中断杂交作图(interrupted mating experiment) 147
中国水仙(*Narcissus tazetta L.* var. *chinensis Roem*) 685
中间缺失(intertitial deletion) 667
中期Ⅰ 30
中心法则(central dogma) 29
中心区 Y 543
中性突变(neutral mutation) 641
中性突变学说(neuture-mutation) 828
中性型小菌落(neutral petite) [P-N]628
中央成分(central element) 707
终变期(diakinesis) 30
终止密码子(termination codons or stop codons) 367,372,418
终止区域(ter D,A 和 ter C,B) 271
终止位点 271
终止子(terminator,t) 304,444,483
种系突变(germ-line mutations) 640
种质(germ plasm) 2
种质论(germplasm theory) 2
重度联合免疫缺损 SCID(severe combined immunodeficint) 556
重链(Heary chain) 240,551
周期蛋白(cyclin) 280,291,580,589
周期蛋白 CDK(cyclin-CDK) 597
周质(periplasmic space) 433
轴向分子(axial element) 707

侏儒(先天性软骨发育不全) 10
珠蛋白 128
珠蛋白链 851
珠孔(micropyle) 35
助病毒(helper virus) 751
转导(transduction) 135,158
转导病毒(transducing viruses) 751
转化(tranformation) 135, 165,592, 594
转化基因(tranforming gene) 594
转化基因组(transforming genome) 751
转化灶(focus) 594
转换(transition) 640
转接器(adaptor) 375
转录(transcription) 293
转录后的加工(posttranscriptional modification) 321
转录激活(transcriptional activation) 255
转录间隔区(TS) 227
转录因子 TF ⅢA 576
转染(transfection) 135,533
转位反应 414
转移(metastasis) 592
转移链 731
转移区(transfer region) 141
转运 RNA ,tRNA 375
转酯反应(transesterification) 339
转酯模型 ,358
转座(transposition) 718,720
转座的 CACTA 群 739
转座酶(transposase) 720,723
转座模型 729
转座因子(transposable element) 676,718
转座重组 698
转座子(transposon, Tn) 723
转座子 Tn3 723
装配因子(assembly factor) 318
椎实螺(*Limnaea*) 633
准性生殖系统(parasexual systems) 93
着色性干皮病(Xeroderma pigmentosum) 658
着丝粒(Kinetochore) 204
着丝粒融合 676
着丝粒作图(centromere mapping) 85
滋养细胞 531,774
子囊(ascas) 83
子囊孢子(ascospore) 83
子实体 83
子细胞 24
姊妹染色体单体差别染色体(sisterchromatid differential staining,SCD) 699
紫茉莉(*Mirabilis jalapa*) 62,623
紫外线(Ultraviolet light rays,UV) 646
自发倒位 673
自发突变(spontaneous mutatiow) 640
自缀容貌综合征 123
自然选择(natural selection) 840
自然转化(natural transformation) 135
自私 DNA(selfish DNA) 223
自由组合定律(law of independent assortment) 13
自主发育(autonomous development) 116
自主复制顺序(autonomously replicating sequences) 288
自主剪接(self-splicing or autosplicing) 339
自主性因子(antonomous elements) 737
阻遏(repression)效应 448
阻遏建立(repressor establishment) 521
阻遏物 CⅠ蛋白 504
组成型基因(constitutive genes) 442
组成型突变(constitutive mutants) 451
组成型异染色质 (constitutive heterochromatin) 33 ,195,563
组蛋白 195,565
组合模型假设 793
组合事件(combining probability rule) 16
组织相容性抗原(major histocom-patibility,MHC) 66
座位控制区(LCR, locus control regiou) 568

A1C018613379